

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2004, 03:08:33 ; Search time 498.358 Seconds
(without alignments)
9692.234 Million cell updates/sec

Title: US-10-658-691-1

Perfect score: 1137

Sequence: 1 gtggtaaccaaccatcaat.....gcattcaggagcacgttga 1137

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	419.4	36.9	1161	2 AAV68073	AAV68073 DNA encod
2	419.4	36.9	1200	2 AAV68074	AAV68074 DNA encod
3	49	4.3	2000	7 ADA71938	ADA71938 Rice gene
4	46.2	4.1	3851	7 ACA40918	ACA40918 Prokaryot
5	46.2	4.1	5496	7 ACA38681	ACA38681 Prokaryot
6	46.2	4.1	6564	4 AAF31637	AAF31637 Mycobacte
7	46.2	4.1	110000	4 AAI99682_32	Continuation (33 o
8	46.2	4.1	110000	4 AAI99683_32	Continuation (33 o
9	43.6	3.8	536	9 ADB68842_Minority	ADB68842 Minority
10	42.8	3.8	2000	7 ADA71938	ADA71938 Rice gene
11	40.6	3.6	2823	7 ACA38373	ACA38373 Prokaryot
12	40.6	3.6	2826	7 ACA40558	ACA40558 Prokaryot
13	40.6	3.6	110000	4 AAI99682_20	Continuation (21 o
14	40.6	3.6	110000	4 AAI99683_20	Continuation (21 o
15	40.2	3.5	1830	7 ACA37565	ACA37565 Prokaryot
16	39.8	3.5	109519	5 AAS08693	AAS08693 Micromono
17	39.2	3.4	1308	7 ACA38683	ACA38683 Prokaryot
18	39.2	3.4	2703	7 ACA40727	ACA40727 Prokaryot
19	39.2	3.4	110000	4 AAI99682_31	Continuation (32 o
20	39.2	3.4	110000	4 AAI99683_31	Continuation (32 o
21	39	3.4	536	9 ADB68842_Minority	ADB68842 Minority
22	38.6	3.4	690	5 AAH67856	AAH67856 C glutami
23	38.6	3.4	813	4 AAF67978	AAF67978 Corynebac

C	24	38.6	3.4	349980	5	AAH68533	Aah68533 C glutami
	25	38	3.3	840	7	ACA45847	Aca45847 Prokaryot
	26	37.6	3.3	1581	7	ACA27336	Aca27336 Prokaryot
C	27	37.4	3.3	2838	2	AAQ41087	Aaq41087 Insulinom
	28	37.4	3.3	2838	5	AAQ4567	Aad04567 Human ins
C	29	37.4	3.3	2838	6	ABL5833	Ab165833 Lung canc
	30	37.4	3.3	2838	6	ABL5856	Ab165856 Lung canc
C	31	37.4	3.3	2838	6	ABL5390	Ab165390 Lung canc
	32	37.2	3.3	3176	5	AAH81799	Aah81799 Human dif
	33	37.2	3.3	3226	3	AZ65082	Aaz65082 Membrane-
	34	37.2	3.3	3226	3	AAH77665	Aaa77665 Human PRO
	35	37.2	3.3	3226	4	AAH21475	Aas21475 Human CDN
	36	37.2	3.3	3226	5	AAH44228	Aaf44228 Human PRO
	37	37.2	3.3	3226	7	ABX77940	Abx77940 Human PRO
	38	37.2	3.3	3226	7	ABX80352	Abx80352 Novel hum
	39	37.2	3.3	3226	7	ACA69258	Aca69258 Human CDN
	40	37.2	3.3	3226	7	ACD24084	Acd24084 Novel hum
	41	37.2	3.3	3226	7	ABX90329	Abx90329 Human sec
	42	37.2	3.3	3226	7	ABX64175	Abx64175 cDNA enco
	43	37.2	3.3	3226	7	ACA67225	Aca67225 cDNA enco
	44	37.2	3.3	3226	7	ACA64397	Aca64397 Novel hum
	45	37.2	3.3	3226	7	ACA03834	Aca03834 cDNA enco

ALIGNMENTS

RESULT 1
AAV68073
ID AAV68073 standard; DNA; 1161 BP.
XX AC AAV68073;
XX AC AAV68073;
DT 12-JAN-1999 (first entry)
DE DNA encoding an oxidase.
XX DE
KW Oxidase; dioxin removal; ds.
XX OS Pseudomonas sp.
XX OS
FH Key Location/Qualifiers
FT CDS 1..1161
FT /tag= a
FT /product= "oxidase"
XX
PN JP10257895-A.
XX
PD 29-SEP-1998.
XX
PF 18-MAR-1997; 97JP-00084401.
XX
PR 18-MAR-1997; 97JP-00084401.
XX (ASAH) ASAH KASEI KOGYO KK.
PA Aaf31637 Mycobacte
XX Continuation (33 o
DR Continuation (33 o
XX Continuation (33 o
XX ADB68842_Minority
PT ADA71938 Rice gene
PT ACA38373 Prokaryot
XX ACA40558 Prokaryot
XX Continuation (21 o
PS Continuation (21 o
XX ACA37565 Prokaryot
CC AAS08693 Micromono
CC ACA38683 Prokaryot
CC ACA40727 Prokaryot
XX Continuation (32 o
SQ Continuation (32 o
Aah67856 C glutami
Aaf67978 Corynebac

Query Match 36.9%; Score 419.4; DB 2; Length 1161;
Best Local Similarity 62.6%; Pred. No. 6.9e-112;

		Matches	689;	Conservative	0;	Mismatches	406;	Indels	6;	Gaps	2;
QY	31	AGAACCAAGGTTTGGAGGCTTATATCGGTGCAAACTCGGTTCCGAAACCAATTGGTAT	90								
Db	37	AGAGTAAAGGCTGGGCGCCCTACGTGGATGGAAGTAGGCTTTCGCAATCAITGGTAC	96								
QY	91	CCGTTTCGCTCGCAGCGAAATCGCCGAAGTACTCCGGTCCCGTCAAGCTCTCTGGGA	150								
Db	97	CCGTCATGTTTTCGAAAGAGATCGACGAGGCGGAGCGAAGACACTAAAACTGCTCGGT	156								
QY	151	GAGAAGATTCTCTCAATCGCTGGCGCGCAAGGCTTATCGATCCAGGACAGGTGCCTG	210								
Db	157	GAGAACTTCTCTCATCTGATATCGATGGGAAGCTGTATTCCTCAAGGACCGTCCCTG	216								
QY	211	CATCGGCTGTAACGCTTTCGACCGGCTGAGTCTATTCGAGAAACACCATATCTCTGC	270								
Db	217	CATCGCGGCTCGAGTTGCTGCTCAAGTTCGAGTCAAAACGAAAGTGCAGCATCACATGC	276								
QY	271	TGGTATCAGGCTGGACATATCGCTGGGACGATGCGCGCTCGTATCTCTCAAAAC	330								
Db	277	TGGTATCAGGCTGGACATATCGCTGGGACGATGCGCGCTCGTATCTCTCAAAAC	336								
QY	331	CCCGGAGTGTGAGATCGCGCGCGCTTTTGAAGAGCTTCCCGTTTGAAGAGCCAAA	390								
Db	337	CCGACAGCGCAGATCGTTCGACAAAGCTGAAACTTACCAGTGCAGGAGCCAAAG	396								
QY	391	GGTCTTATCTGTTTACGTAGCGACGCGGACCAACCGCGCTTATCGAAGATGTCGG	450								
Db	397	GGTCTGCTCTTCAATTTATCTTGGGATGGGACCGCTTCTCCCTTGGCGCGATAGCGCA	456								
QY	451	CCCGGCTCTCTGATGAAACCGCGCCATTTCAGCGCAACATCGGCTCGTGGCTCGAAC	510								
Db	457	CCCAATTTCTTGACATGACATGGAATCTCTCGGAGAAACCAATCATCAAGTCTAAC	516								
QY	511	TGGCGCTTGGTCCGAAACCGCTTTGATGGGGACGCTTTCATTTCAGAGATTCG	570								
Db	517	TGGCGCTTGGTCCGAAACCGCTTTGATGGGGACGCTTTCATTTCAGAGATTCG	576								
QY	571	ATCCTGCTGAAGGCAACGATATCATCTTCCGCTTGGCTTTTCCGCC---TGGCGATGCC	627								
Db	577	ATTCCTGCTGAAGGCAACGATATCATCTTCCGCTTGGCTTTTCCGCC---TGGCGATGCC	636								
QY	628	GACAGCTTACGGT---TCCGAGTTGCTGGGGCAAGCCCAAGGTTTACGATCTG	684								
Db	637	AAGCAACAACTCGTGTGTTGACGATGACGCTGTCGAGCGAAGGTTTACGATCTT	696								
QY	685	CTTGGCGAGATTCGGTCCGCTTTTTCGAGGAGCATGATCGAAGGCAACCTCAATCCAT	744								
Db	697	ATTGGCGAACATCGGCTCCAGTGTGAGGNACTATCGGGGCAAGTGTCCGCGNA	756								
QY	745	GGCAACATTGGCAGCAAGCGGCTGCGCATCAGCATATCGATCTGGTCCGCGGCTACTC	804								
Db	757	GGTGGCTACGGCGAAATAATTGAGCGAACGATATCTCCATTTGGCTCCCGGCTGTCTC	816								
QY	805	AGGTCGAAACCGTGGCGGATCCGAGCTCACGAGTTCGAATGTTACGTCGGTTCGAT	864								
Db	817	AGGTCGAAACCGTGGCGGATCCGAGCTCACGAGTTCGAATGTTACGTCGGTTCGAT	876								
QY	865	GAGACGAGCACCCTCTACTTCCAGAGCTGGGCAAAAGTCTGTGACGTCGAAAGGAGCGCA	924								
Db	877	GAAACACACACTATTACTTCCAAACTCTTTGGCAAAACCATGTGCCAATGACGAGGAACG	936								
QY	925	GATCTCTTCGAGCGAGATTCACGAAATGAGTGGTAGGCTCGCGTTAACGCTTCAT	984								
Db	937	ARGAATTTACGAAACAGAGTTTCGAAAGCAAGTGGAAACCGATGGCGCTCGAAGGATTCAC	996								
QY	985	GATGACGACATCATGGCAGCTGAATGATGGAGCCGTTCTTACGCTGATGATCGGTTGG	1044								
Db	997	AACGATGACATCTGGGCTCGGAGCTATGTTGATTTCTTACGCGCATGATGAAGCTGG	1056								
QY	1045	TCCGAGAAATCTCTTTCGAGCGGACCGCGCAATCATCGAGTGGCGGGCTTGGCAGT	1104								
Db	1057	GTCAACGAGATTTTGTTCGAGGTGGACGAGGCTATCGTGGCATGGCGCAAGCTGGCGAGC	1116								

QY	1105	CAGCAATCGCGCATTCAG	1125
Db	1117	GAACACAATCAGGTTATTCAG	1137
RESULT 2			
AAV68074			
ID	AAV68074 standard; DNA; 1200 BP.		
XX	AAV68074;		
XX	12-JAN-1999 (first entry)		
XX	DNA encoding an oxidase.		
XX	Oxidase; dioxin removal; ds.		
XX	Pseudomonas sp.		
XX	Key	Location/Qualifiers	
XX	CDS	1..1200	
XX	FT	/*tag= a	
XX	FT	/product= "oxidase"	
XX	FT	/note= "no termination codon given"	
XX	JPI0257895-A.		
XX	29-SEP-1998.		
XX	18-MAR-1997;	97JP-00084401.	
XX	18-MAR-1997;	97JP-00084401.	
XX	(ASAH) ASahi Kasei Kogyo KX.		
XX	WPI; 1998-575906/49.		
XX	P-PSDB; AAW80332.		
XX	Oxidase gene derived from a microbe for removal of hetero:poly:cyclic aromatic hydrocarbon to a benzol		
XX	Claim 2; Page 7-8; 15pp; Japanese.		
XX	The present sequence encodes an oxidase which is de		
XX	and converts a heteropolycyclic aromatic hydrocarb		
XX	homologue. The oxidase is used in a method for the		
XX	which dibenzo-p-dioxin is converted to a diphenyl e		
XX	Sequence 1200 BP; 323 A; 284 C; 338 G; 255 T; 0 U; 0		
QY	31	AGAACCAAGGTTTGGAGGCTTATATCCGTGCGAAACTCGGG	90
Db	94	AGAGTAAAGGCTGGGCGCCCTACGTGGATGCGAAGCTAGG	153
QY	91	CCCGTTCCGCTCGCAGCGAAATCGCGAAGTACTCCCGTT	150
Db	154	CCGTTTCGCTCGCAGCGAAATCGCGAAGTACTCCCGTT	153
QY	151	GAGAGATTCGCTCAATCCGCTGGGCGCAAGTCTATGCG	210
Db	214	GAGAACTTTCGCTCGTCAATCGTATCGATGCGAAGCTG	273
QY	211	CATCGGCTGTAAAGCTTTTCGACCGGCTCGAGTCTATT	270
Db	274	CATCGGCGCTCGAGTTGTCGTTCAAGTTCAGTTCGCAAA	333
QY	271	TGGTATCAGCGCTGGACATATCGCTGGGACGATGCGCG	330

DB 334 CYRWRKWKTKYKRCYVATYCCYKRGWYSRRSMETAGKWMRSWNR 275
QY 545 GCACGCTCTTACACAGATTCGATCCCTGGTGAAGGCAACGATATCTGCGC 604
DB 274 SYSWYMYKWKYKYSYMSYWARSGTWSRGAARTYKYSTSRKWMRACRMYACR 215
QY 605 TTGGCTTTGCGCTGCGATCCGACGCTTACGCTTCCGAGTTCTCGGGCAAGC 664
DB 214 RYRSTYYCGSYGSSWKYMSKSMRMTCSNCSCYCYCYGAMCWSGMSMYNGS 155
QY 665 CCAAGGCTTTACGATCTGC 685
DB 154 CGCTYRGWKRYSKNCCKY 134

RESULT 4
ID ACA40918 standard; DNA; 3851 BP.
XX ACA40918;
XX
DT 19-JUN-2003 (first entry)
DE Prokaryotic essential gene #22575.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX Mycobacterium tuberculosis.
XX PN WO20027183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US0009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WFI; 2003-029926/02.
XX P-PSDB; ABU37048.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 28788; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIFO at
XX ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 3851 BP; 647 A; 1292 C; 1274 G; 635 T; 0 U; 3 Other;
Query Match 4.1%; Score 46.2; DB 7; Length 3851;
Best Local Similarity 46.9%; Pred. No. 0.011;
Matches 144; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
QY 614 CGCTGGCGATCCGACCACTTACGCTTCCGAGTTGTCGGGCAAGCCCAAGGTG 673
DB 1009 CGCTGGAAGAGACCGCGAGCTACGGTGCCTGTGTGCTGATCGGTGACCCGCGCG 1068
QY 674 TTACGATCTGCTGGCGAGCATTCGCTGCGGGTTTCGAAGCATGATCGAAGGCAAC 733
DB 1069 GTGACCGCTGCTGCCCGCGCTGCACTGCTGCGGACCGCATCAGCACTCGAGC 1128
QY 734 CTGCAATCCATGGCAACATTGGCAGCAAGCGCTGCGCATCAGCATATCGATCTGGCTGC 793
DB 1129 TGGCATCCGCGCGGCGAGGACGCTGCGCGCGGATCAGCCGAACTCTCTGGGACG 1188
QY 794 CGGCGCTACTCAAGTTCGACCGTGGCGGATCCGAGCTCAGCGAGTTCGATGCTACG 853
DB 1189 GGATCGTGTGTGTCTGCTCCGCCCGAGCAACGATCGATGCCGACGAGCTCAAC 1248
QY 854 TGCAGTTCGATGAGACGAGCAACCTCTACTTCCAGACGCTGGGCAAGTCTGACGCTCA 913
DB 1249 TGGAGTTGGCAGCAGCAGCAGCGCTGCTGCTGCCAGCGTGGTTCGAGACGCTGACGGA 1308
QY 914 AGGAAGC 920
DB 1309 TGGGTGC 1315

RESULT 5
ID ACA38681 standard; DNA; 5496 BP.
XX ACA38681;
XX AC ACA38681;
XX DT 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #20338.
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX OS Mycobacterium bovis.
XX PN WO20027183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US0009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX DR P-PSDB; ABU34811.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 14; SEQ ID NO 26551; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
XX CC prokaryotic essential genes. Note: the sequence data for this patent did
XX CC not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 5496 BP; 941 A; 1859 C; 1795 G; 901 T; 0 U; 0 Other;
Query Match 4.1%; Score 46.2; DB 7; Length 5496;
Best Local Similarity 46.9%; Pred. No. 0.013;
Matches 144; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 614 CGCTGGCGATCCCGACGCTTACGGTTCCGAGGTGCTCGGGCAAGCCCAAGGTG 673
Db 2648 CGCTGGAAGAAGACCGCGAGCTAGGGTGCCCTGTGCTGATCGGTGACCCCGCGG 2707
QY 674 TTACGATCTCTTTGGCAGCATTCGGTGGCGGTTTTTCGAAGGCATGATCAAGGCAAC 733
Db 2708 GTGACCCGCTGCTGCCCGCGCTGACGTGCTGCTGCGCGACCGCATCACGACCTCGAGC 2767
QY 734 CTGCAATCCATGGCAACATTGGCGAGCGGTGCGCATCAGCATATCGATCTGGCTGC 793
Db 2768 TGGCATCCGCGCGGCAAGACCCAGCTGCGCGCGGGGATCAGCCGACCTCTTGGGACG 2827
QY 794 CGGGCGTACTCAAGTTCGAACCGTGGCGGATCCGAGCTCACCGAGTTGGAATGGTACG 853
Db 2828 GGATCGTTGTGCTGCTCCGCCCGGAGCGAAGACGAAATCGATCCGACGAGGCTCAAC 2887
QY 854 TGCGGTGATGAGACGACCACTTACTTCCAGACGCTGGGCAAGATGCTGACGTCAA 913
Db 2888 TGGAGTTGGCACGCAACACGACGCTGCTGGTCCGACGCTGCTGACACCGTACGCGAA 2947

QY 914 AGGAAGC 920
Db 2948 TGGGTGC 2954

RESULT 6
AAF31637
ID AAF31637 standard; DNA; 6564 BP.
XX AAF31637;
XX AC AAF31637;
XX DT 09-APR-2001 (first entry)
XX DE Mycobacterium tuberculosis ppsC gene.
XX KW Mycobacterium tuberculosis; attenuated microorganism;
XX KW signature tagged transposon mutant; mutant library;
XX KW mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
XX KW vaccine; ppsC; ds.
XX OS Mycobacterium tuberculosis.
XX FN WO200102555-A1.
XX PD 11-JAN-2001.
XX PF 06-JUL-2000; 2000WO-IB000950.
XX PR 06-JUL-1999; 9US-0142982P.
XX PR 08-JUL-1999; 99US-0142833P.
XX PA (INSP) INST PASTEUR.
XX PI Gicquel B, Guilhot C, Camacho L;
XX DR WPI; 2001-091804/10.
XX DR P-PSDB; AAB66467.
XX PT Screening a mutant library for mutants unable to grow under specific
XX PT conditions and for identifying loci involved in pathogenicity, comprises
XX PT using signature tagged transposon mutagenesis.
XX PS Example 8; Page 122-124; 159pp; English.
XX CC The present sequence is given in a specification relating to a method for
XX CC screening a library of mutants. The method comprises constructing a
XX CC library with insertions in genes and/or regulatory regions of the
XX CC organisms of interest, where the insertion contains a tag and/or a
XX CC transposon associated with a tag. The mutants are identified by
XX CC hybridisation of the tags to known sequences. The method is useful for
XX CC treating an individual suffering from a mycobacterial infection,
XX CC suspected of being infected with a Mycobacterium, or having been exposed
XX CC to an infectious Mycobacterium. It is also useful for identifying and
XX CC isolating mutants of actinomycetales and for identifying compounds that
XX CC have antibiotic activity. The method is used to identify mutants of
XX CC microorganisms, preferably an actinomycetales, such as *M. tuberculosis*,
XX CC *M. bovis*, *M. leprae*, *M. avium*, *M. intracellulare* and *M.*
XX CC *paratuberculosis*, that is unable to grow under specific conditions. It is
XX CC especially useful for identifying loci involved in pathogenicity. It is
XX CC useful in constructing vaccines. The method can be used to screen
XX CC multiple libraries concurrently. It can screen libraries of different
XX CC organisms or different strains of the same organism
XX SQ Sequence 6564 BP; 1137 A; 2189 C; 2154 G; 1084 T; 0 U; 0 Other;
Query Match 4.1%; Score 46.2; DB 4; Length 6564;
Best Local Similarity 46.9%; Pred. No. 0.014;
Matches 144; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 614 CGCTGGCGATCCCGACGCTTACGGTTCCGAGGTGCTCGGGCAAGCCCAAGGTG 673
Db 3716 CGCTGGAAGAAGACCGCGAGGCTAGGGTGCCCTGTGCTGATCGGTGACCCCGCGG 3775

QY 674 TTTACGATCTCTGGGAGCATTCGGTGGCGGTTTTTCGAGGAGCATGATCGAAGGCAAC 733
Db 3776 GTGACCCGCTGCTGCCCGGCTGCAGTCTGCTGCGGACCGCATCACCGACTCGAGC 3835
QY 734 CTGCAATCCATGGCAACATTTGGGAGCAAGCGCTGCGCATCAGCATATCGATCTGGCTGC 793
Db 3836 TGGCATCGCGCGCGACCAAGCCAGCTGCGCGCGCGGATCAGCGCAACCTCTCGGACG 3895
QY 794 CGGGGTACTCAAGGTGCAACCGTGGCGGATCCGAGCTCACGAGTTCGAATGGTACG 853
Db 3896 GATCGTTGCTGTGTCGCGCGGAGCAAGCAGCATCATCGCGGAGGCTCAAC 3955
QY 854 TCGCGGTGATGAGACGACCTCTTACTTCCAGACGCTGGGAAAGTGTGACGTCAA 913
Db 3956 TGGAGTTGGCAGCACACCGAGCTGTGCTGCGCAGCGTGGTCGAGACCGTGCACGGA 4015
QY 914 AGGAAGC 920
Db 4016 TGGGTGC 4022

RESULT 7
AAI99682_32
Continuation (33 of 45) of AAI99682 from base 3200001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529

Query Match 4.1%; Score 46.2; DB 4; Length 110000;

Best Local Similarity 46.9%; Pred. No. 0.043;
Matches 144; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
QY 614 CGCTGGCGATCCGACAGCTTACGCTTCGAGTTCGCTGGGCAAGCCCAAGGTG 673
Db 59397 CGCTGAAAAGACCGCGAGGCTACGGGTGCCCTTCTGCTGATCGGTGACCCCGCGG 59456
QY 674 TTTACGATCTGTGTGGCGAGCATTCGGTGGCGGTTTTTCGAGGAGCATGATCGAAGCAAC 733
Db 59457 GTGACCCGCTGTGCGCGCTGCACTGCTGCTGCGGACCGCATCACCGACCTCGAGC 59516
QY 734 CTGCAATCCATGGCAACATTTGGCAGCAAGCGGTCGCCATCAGCATATCGATCTGGCTGC 793
Db 59517 TGGCATCCCGCGCGGAGCAAGCAAGCTGCGCGCGCGGATCAGCGCAACCTCTGGGAGC 59576
QY 794 CGGGGTACTCAAGGTGCAACCGTGGCGGATCCGAGCTCACGAGTTCGAATGGTACG 853
Db 59577 GATCGTTGCTGTGTCGCGCGGAGCAAGCAGCATCATCGCGGAGGCTCAAC 59636
QY 854 TCGCGGTGATGAGACGACCTCTTACTTCCAGACGCTGGGCAAGTGTGACGTCAA 913
Db 59637 TGGAGTTGGCAGCACACCGAGCTGTGCTGCGCAGCGTGGTCGAGACCGTGCACGGA 59696
QY 914 AGGAAGC 920
Db 59697 TGGGTGC 59703

RESULT 8
AAI99683_32
Continuation (33 of 44) of AAI99683 from base 3200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000

PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX
XX Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
XX Claim 27; SEQ ID NO 5263; 899pp; English.
XX
XX The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
SQ
Query Match 3.8%; Score 42.8; DB 7; Length 2000;
Best Local Similarity 8.8%; Pred. No. 0.084;
Matches 61; Conservative 329; Mismatches 301; Indels 5; Gaps 3;
QY 85 TGGTATCCGCTTCCGCTCGAGCGAAATCGCGAAGTACTCCGCTCCGCTCAAGTTC 144
DB 15 WRGRRYVWAGMWSRCARMSRMSRMSRMSRMSRMSRMSRMSRMSRMSRMSRMSR 74
QY 145 CTGGAGAGAAATTCGTCAATCGGTGGCGGCAAGTCTATCGGATCCAGACAGG 204
DB 75 RTGSKWSGSGSYGKMKYKRSKRWGRGRGRGRGRGRGRGRGRGRGRGRGRGRGR 134
QY 205 TGCTCATCGCGGTGTAAGCTTTCGACCGGTGAGTCTATTCCAGAACACCAT 264
DB 135 MGGGRMSYVWMCYARGCGSKRKSKGSGMKTCTRRGARGGSGWSGAKYSGMSKR 194
QY 265 TCTGTGTGATCAGCGTGGACATATCGTGGAGCATGGCGGCTCGTATCGTATC 324
DB 195 MMSSCGRSGCGRRSAYRYGTSRYKTYKTYKTYKTYKTYKTYKTYKTYKTYKTY 254
QY 325 ACAACCCCGGAGTGCAGATCGCGCGCGCTTTGAGACGCTCCCGTTGAGAG 384
DB 255 RSKRRSMWMMKMRMSRSGYSGYSGYSGYSGYSGYSGYSGYSGYSGYSGYSGY 314
QY 385 GCCAAAGGTCTTATCTTCTGTAC--GTAGCGCAGCGCGAAACCAACGCGCTTATCGAAG 442
DB 315 GRGYMSRMAMMYKMYRYGKMGKMGWAGRMWMSRMSRMSRMSRMSRMSRMSR 374
QY 443 ATGTGCGGCGCGCTTCTGTGATAAACCAGCGCATTCACGCG--CAACATCGGCTGTG 501
DB 375 WAKKSRTSRKKRKMCRKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 434
QY 502 GCTCTGAACGCGCTTGGGTGCGGAAACGCGTGTGTCGCGGCGACGCTTCTATTCAC 561
DB 435 RMKSYGMRMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMS 494
QY 562 AAGAATTCGATCTGTTGAGGCGCAAGATATCATCTCGCGCTTGGCTTGGCGCTGCG 621
DB 495 GMRKCRRRRWRGMYRMWRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 554
QY 622 GATCCGACAGCTTACGCTTCCGAGTTGCTGCGGCA--AGCCCAAGGTGTTTACG 679
DB 555 WMKRYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYK 614
QY 680 ATTCTGCTGCGAGCATTCGCTCGCGTGTTCGAGAGCATGATCGAGCAACACCTGCAA 739
DB 615 RSSRKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKR 674
QY 740 TCCATGSCACATTTGGCAGCAGCGCGTCCCATCA 775

Db 675 TYAKGYSWRYRWYRWYRWYRWYRWYRWYRWYRWYRWYRWYRWYRWYRWYRWY 710

RESULT 11

ACA38373
ID ACA38373 standard; DNA; 2823 BP.
XX
XX ACA38373;
XX
XX 19-JUN-2003 (first entry)
DT
DE Prokaryotic essential gene #20030.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
XX Mycobacterium bovis.
OS
XX WO200277183-A2.
PN
XX
XX 03-OCT-2002.
PD
XX
XX 21-MAR-2002; 2002WO-US009107.
PF
XX
XX 21-MAR-2001; 2001US-00815242.
PR
XX 06-SEP-2001; 2001US-00948993.
PR
XX 25-OCT-2001; 2001US-0342923P.
PR
XX 08-FEB-2002; 2002US-00072851.
PR
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI
XX
XX WPI: 2003-029926/02.
DR
XX P-PSDB; ABU34503.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 14; SEQ ID NO 26243; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did

CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2823 BP; 478 A; 949 C; 921 G; 475 T; 0 U; 0 Other;

Query Match 3.6%; Score 40.6; DB 7; Length 2823;
Best Local Similarity 49.8%; Pred. No. 0.42;
Matches 103; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 346 ATCGCGCGCGCTTTGAAGACGTTCCGCTTGAAGGCGCAAGCTCTATCTTCGTT 405
Db 667 AGCGCGCGATACGACTGCTGCTGCTGCAAGCGCCACGACCGGCGACTG 726
QY 406 TACGTAGCGCGGCGCAACCGCGCTTATCGAAGATGTCGCGCGGCTTCCTTGAT 465
Db 727 GTGGCGGTCGCGCGGCGACTTGTGGCGTACGCTGCGCGCGCGAGATCGGC 786
QY 466 GAAACCGCGCATTCACGGCAACATCGCTGCTGCGCTGCAACTGCGCTGGGTCG 525
Db 787 GCTGACGTCGCTTGGCACCACACACACGCTTCGAGTGCAGTGGGTTTGGCGGCG 846
QY 526 GAAACCGCTTTGATGCGGCGCACGTC 552
Db 847 CATCGCGGTACCTTGGGTGACGCGC 873

RESULT 12
ACA40558
ID ACA40558 standard; DNA; 2826 BP.
XX
AC ACA40558;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #22215.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
OS Mycobacterium tuberculosis.
XX
FN WO20027183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu RH;
XX
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU36688.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 28428; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway; (8)
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2826 BP; 481 A; 949 C; 920 G; 476 T; 0 U; 0 Other;

Query Match 3.6%; Score 40.6; DB 7; Length 2826;
Best Local Similarity 49.8%; Pred. No. 0.42;
Matches 103; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 346 ATCGCGCGCGCTTTGAAGACGTTCCGCTTGAAGGCGCAAGCTCTATCTTCGTT 405
Db 667 AGCGCGCGATACGACTGCTGCTGCTGCAAGCGCCACGACCGGCGACTG 726
QY 406 TACGTAGCGCGGCGCAACCGCGCTTATCGAAGATGTCGCGCGGCTTCCTTGAT 465
Db 727 GTGGCGGTCGCGCGGCGACTTGTGGCGTACGCTGCGCGCGAGATCGGC 786
QY 466 GAAACCGCGCATTCACGGCAACATCGCTGCTGCGCTGCAACTGCGCTGGGTCG 525
Db 787 GCTGACGTCGCTTGGCACCACACACACGCTTCGAGTGCAGTGGGTTTGGCGGCG 846
QY 526 GAAACCGCTTTGATGCGGCGCACGTC 552
Db 847 CATCGCGGTACCTTGGGTGACGCGC 873

RESULT 13
AAI99682.20
Continuation (21 of 45) of AAI99682 from base 2000001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 100001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000

Search completed: June 19, 2004, 04:29:05
Job time : 504.358 secs

XX		(ELIT-) ELITRA PHARM INC.	
PA			
XX		Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;	
PI		Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;	
XX			
DR		WPI; 2003-029926/02.	
XX		P-PSDB; ABU33695.	
XX			
PT		New antisense nucleic acids, useful for identifying proteins or screening	
PT		for homologous nucleic acids required for cellular proliferation to	
PT		isolate candidate molecules for rational drug discovery programs.	
XX			
PS		Claim 14; SEQ ID NO 25435; 1766bp; English.	
XX			
CC		The invention relates to an isolated nucleic acid comprising any one of	
CC		the 6213 antisense sequences given in the specification where expression	
CC		of the nucleic acid inhibits proliferation of a cell. Also included are:	
CC		(1) a vector comprising a promoter operably linked to the nucleic acid;	
CC		encoding a polypeptide whose expression is inhibited by the antisense	
CC		nucleic acid; (2) a host cell containing the vector; (3) an isolated	
CC		polypeptide or its fragment whose expression is inhibited by the	
CC		antisense nucleic acid; (4) an antibody capable of specifically binding	
CC		the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular	
CC		proliferation or the activity of a gene in an operon required for	
CC		proliferation; (7) identifying a compound that influences the activity of	
CC		the gene product or that has an activity against a biological pathway	
CC		required for proliferation, or that inhibits cellular proliferation; (8)	
CC		identifying a gene required for cellular proliferation or the biological	
CC		pathway in which a proliferation-required gene or its gene product lies	
CC		or a gene on which the test compound that inhibits proliferation of an	
CC		organism acts; (9) manufacturing an antibiotic; (10) profiling a	
CC		compound's activity; (11) a culture comprising strains in which the gene	
CC		product is overexpressed or underexpressed; (12) determining the extent	
CC		to which each of the strains is present in a culture or collection of	
CC		strains; or (13) identifying the target of a compound that inhibits the	
CC		proliferation of an organism. The antisense nucleic acids are useful for	
CC		identifying proteins or screening for homologous nucleic acids required	
CC		for cellular proliferation or screening for homologous nucleic acids required	
CC		for cellular proliferation to isolate candidate molecules for rational	
CC		drug discovery programs, or for screening homologous nucleic acids	
CC		required for proliferation in cells other than S. aureus, S. typhimurium,	
CC		K. pneumoniae or P. aeruginosa. The present sequence is one of the target	
CC		prokaryotic essential genes. Note: The sequence data for this patent did	
CC		not form part of the printed specification, but was obtained in	
CC		electronic format directly from WIPO at	
CC		ftp.wipo.int/pub/published_pct_sequences	
XX			
SQ		Sequence 1830 BP; 338 A; 591 C; 655 G; 246 T; 0 U; 0 Other;	
	Query Match	3.5%; Score 40.2; DB 7; Length 1830;	
	Best Local Similarity	47.4%; Pred. No. 0.47;	
	Matches 120; Conservative	0; Mismatches 133; Indels 0; Gaps 0;	
Qy	688	GCGAGCATTCGGTCCGCGTTTCGAGGCATGATCGMAGCAACCTGCATCCTATGCCG 747	
Db	1306	GACGAGTGCGCGCTCGGGGTTCATCGACCGCGGTGTGCGTGCATCCGAACAACGTC 1365	
Qy	748	AACATTGCGACAAAGCGGTGCGCCATCAGCATATCGATCTGGCTGCGGGGTACTCAAG 807	
Db	1366	AACCTGGCGTGGCGTCCGACGCGGTGATCATCGGCTTCACGTGCGCGCGGCGAAG 1425	
Qy	808	GTCGAACGTCGCGCGATCCCGAGCTACGCAGTTCGAATGGTAGTTCGGTCCGGTTCGATGAG 867	
Db	1426	GCCACCAGCTGGGCCAACCGCGAGGCGGTGAGATCCCGTACTACTCTGGTGATCTACCG 1485	
Qy	868	ACCAGGCACCTCTACTTCCAGACGCTGGGCAAAGTCGTGACGTCAAAGGAAGCGCGAC 927	
Db	1486	CGATACGAGAGATCGAAGAAGCCCTGCGCGGATGTCTCAAGCGATTCACGAGAGAAC 1545	
Qy	928	TCCTTCGAGCGAG 940	
Db	1546	CAGCTGGCGCGCG 1558	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2004, 04:06:17 ; Search time 3321.09 Seconds
(without alignments)
10223.528 Million cell updates/sec

Title: US-10-658-691-1
Perfect score: 1137
Sequence: 1 gggcctaaccacatcaat.....gcatcaggaagcagcttga 1137

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpi:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_red:*
26: em_gss_pmg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	43.4	3.8	771	29	CG368223
C 2	42.2	3.7	1201	13	BX381961
C 3	42.2	3.7	1294	28	BZ550514
C 4	40.4	3.6	493	28	BH872653

C	5	40.4	3.6	626	28	BH872654
C	6	40.4	3.6	841	29	CC731706
C	7	40.4	3.6	869	29	CC731699
C	8	40	3.5	916	12	BG281668
C	9	39.6	3.5	515	13	EX424977
C	10	39	3.4	379	13	BQ632394
C	11	39	3.4	547	14	CB023968
C	12	39	3.4	697	13	CA146289
C	13	39	3.4	761	14	CB028106
C	14	38.8	3.4	645	29	CNS01213
C	15	38.4	3.4	657	9	AJ489090
C	16	38.4	3.4	808	14	CA255857
C	17	38.2	3.4	557	13	BQ236160
C	18	38.2	3.4	622	9	AI637136
C	19	38.2	3.4	729	28	AQ796054
C	20	38	3.3	583	14	CF919632
C	21	38	3.3	611	12	BI387467
C	22	38	3.3	635	14	CA254345
C	23	38	3.3	660	13	CA145013
C	24	38	3.3	709	29	CC695831
C	25	38	3.3	826	29	CG214497
C	26	38	3.3	891	29	CG368338
C	27	38	3.3	916	29	CG323718
C	28	37.8	3.3	449	28	AQ936799
C	29	37.8	3.3	500	9	AU179654
C	30	37.8	3.3	540	28	BZ893211
C	31	37.6	3.3	429	13	EX766306
C	32	37.6	3.3	564	14	CD232653
C	33	37.6	3.3	617	12	BJ246934
C	34	37.6	3.3	851	13	EX627762
C	35	37.6	3.3	1023	12	BI914840
C	36	37.4	3.3	143	10	AW454389
C	37	37.4	3.3	417	13	BQ539085
C	38	37.4	3.3	477	14	CB046467
C	39	37.4	3.3	477	14	CB046471
C	40	37.4	3.3	506	14	CB046470
C	41	37.4	3.3	740	13	EX873986
C	42	37.2	3.3	473	10	AW502888
C	43	37.2	3.3	859	28	BZ575332
C	44	37.2	3.3	988	14	CF884514
C	45	37.2	3.3	1036	9	AL553834

ALIGNMENTS

RESULT 1
CG368223/c
LOCUS CG368223
DEFINITION OG00E64TH ZM.0.7_1.5 KB Zea mays genomic clone ZMMBMa0698K07,
genomic survey sequence.
ACCESSION CG368223
VERSION CG368223.1 GI:34285490
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 771)
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Renick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, N.W., Nurnberg, A., Robbins, D. and Lakey, N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other GSSs: OG00E64TV
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR


```
Class: sheared ends.
Location/Qualifiers
1. 771
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMBM0698K07"
/clone_lib="ZM.0.7.1.5 KB"
/note="Vector: pBSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 3.8%; Score 43.4; DB 29; Length 771;
Best Local Similarity 45.8%; Pred. No. 14;
Matches 149; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 698 CGTGCCGGTTTCGAAGGCATGATCGAAGGCAACCTGCATCATCGACATTTGCA 757
Db 416 CGAGCGGGTTCACCGCCAGCAGCGGAGTGCGCCGCTACGCGACGGCGAGC 357
Qy 758 GCAAGCGCTCGCCATCAGATCATGATCTGGCTGCGGGCGTACTCAAGGTGCAACCGT 817
Db 356 GCGACGGCGCTCAGCTGGACTACGGCTCGCTCGCGACATCGACGCTTCTTCCAGT 297
Qy 818 GCGCGATCCGAGCTACGAGTTCGAATGCTAGTGGCTGCGCTGATGAGACCAACC 877
Db 296 CGCCCAAGTGTAGACGACGACCTACTACTAGCAGCGCTCATGACCCCTCGCGCG 237
Qy 878 TCTACTTCCAGCGCTGGGCAAGTCTGACGTCAAGGAGCGGCGAGACTCTTTCGAGC 937
Db 236 CTTCTTCGCGCGCGCCATCGTTGCGGCGAGCAGCAGCAGCGCGCGCGGAT 177
Qy 938 GAGATTCACGAAATGGTAGGCTCGCGCTTAAAGCTTCAATGATGAGCATCA 997
Db 176 GTGGAGACGCGCGAGCATGCGCTCTGAGCTTCAGCTTCCCGCTCTGGACTGCT 117
Qy 998 TGGCAGCTGAATCGATGGAGCGGTT 1022
Db 116 GAGTGAAGCGCGCAGATTCGCT 92

RESULT 2
BX381961 1201 bp mRNA linear EST 08-MAY-2003
LOCUS BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1072YF05 3-PRIME, mRNA sequence.
ACCESSION BX381961
VERSION BX381961.1 GI:30453007
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1. (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID: CS0D1072CC03NP1.
Location/Qualifiers
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1072YF05"
/tissue_type="PLACENTA COT 25-NORMALIZED"

FEATURES
source
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1072YF05"
/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched double-strand cDNA was
digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 3.7%; Score 42.2; DB 13; Length 1201;
Best Local Similarity 3.4%; Pred. No. 34;
Matches 22; Conservative 238; Mismatches 391; Indels 1; Gaps 1;

Qy 369 GTTCCCGGTTGAAGAGCCAAAGTCTTATCTTCGTTTAGTAGGCGAGCGGCAACAAC 428
Db 415 KKKCKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 474
Qy 429 GCGCTTATCAAGAGTGTGCGCGCGGCTTCCTTGATGAAACCGCGCCATTCACGCCCA 488
Db 475 NNNCKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 534
Qy 489 ACATCGGCTCGTGGCTCGAAC-TGGGGCTTGGGTGCGGAAACGGCTTTGATCGGGGC 547
Db 535 MNAMNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 594
Qy 548 ACGTCTTCATCAAGAATTGATCTGCTGTAAGGCAACGATATCATCTCTCCCGCTG 607
Db 595 MNNNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 654
Qy 608 GCTTGGCGCTGCGATCCCGACGAGTTACGAGCTTCGAGGTTGCTGCGGGCAACCCCA 657
Db 655 NNNNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 714
Qy 668 AAGGTGTTACGATCTGCTTGGCGAGATTCGGTGGCGGTTTTTCGAAGGCGATCATCAAG 727
Db 715 KKKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 774
Qy 728 GCAAACTGCAATCCATGCAACATTTGGCAGCAAGCGCTGCGCATCAGCATATCGATCT 787
Db 775 NNNNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 834
Qy 788 GGCTGCGGGGCTACTCAAGTCAAGCTGCAAGCGTGGCGGATCCGAGCTCAAGCTTCAAT 847
Db 835 MNCCKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 894
Qy 848 GGTACGTGCGCGTGTAGAGCAGCAGCAGCTCTTACTTCCAGAGCTGGGCAAAAGTGTGA 907
Db 895 KNCCKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 954
Qy 908 CGTCAAGGAGGCGGAGACTCTCTGAGCGAGAAATTCACGAGAAATGGGTAGGCTCG 967
Db 955 MKMKCKCKYMBKMKCAKKCKMKYKNCMKYKNCMKYKNCMKYKNCMKYKNCMKYKNC 1014
Qy 968 CGCTTAACGGCTTCAATGATGAGCAGCATCATGCGCAGCTGAATCGATCGAGCC 1019
Db 1015 CAMKVMVMBAMVMBAMVMBAMVMBAMVMBAMVMBAMVMBAMVMBAMVMBAMVMBAMV 1066

RESULT 3
BX550514/c 1294 bp DNA linear GSS 17-DEC-2002
LOCUS BX550514 pacsl-60_2692.y2 pacsl-60 Pseudomonas aeruginosa genomic clone
DEFINITION pacsl-60_2692, genomic survey sequence.
ACCESSION BX550514
VERSION BX550514.1 GI:27154095
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE 1. (bases 1 to 1294)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
```

J. Bacteriol.: (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105
Tel: 20623216954
Tel: 2066857244
Fax: 2066857244
Email: ckraymond@u.washington.
Class: shotgun.

FEATURES	SOURCE
1. High Accuracy: The model achieves a high accuracy rate, consistently performing well across various datasets and scenarios.	Model Performance Metrics
2. Scalability: The system is designed to scale efficiently, handling large volumes of data and complex tasks without significant performance degradation.	System Architecture
3. Interpretability: The model's decisions are transparent and explainable, allowing users to understand the underlying reasoning and identify potential biases.	Model Explainability
4. Robustness: The system is resilient to adversarial attacks and noisy inputs, maintaining high performance under challenging conditions.	Security and Robustness
5. Efficiency: The model is optimized for fast inference times and low resource consumption, making it suitable for real-time applications.	Performance Optimization
6. Flexibility: The system can be easily adapted to different domains and tasks, demonstrating strong generalization capabilities.	Domain Adaptability
7. Integration: The model seamlessly integrates with existing workflows and systems, enabling smooth adoption and deployment.	System Integration
8. Customization: The system offers a high degree of customization, allowing users to tailor the model's behavior to their specific requirements.	User Customization
9. Security: The system implements robust security measures to protect user data and ensure compliance with relevant regulations.	Security Measures
10. Support: Comprehensive documentation and technical support are provided to assist users in maximizing the system's potential.	Customer Support

```

Location/Qualifiers
1..1294
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_2692"
/clone_lib="pacs1-60"
/note="clinical isolate 1-60 Whole library."

```

ORIGIN

Query Match	3.7%;	Score 42.2;	DB 28;	Length 1294;
Best Local Similarity	45.8%;	Pred. No. 35;		
Matches 146;	Conservative 0;	Mismatches 173;	Indels 0;	Gaps 0;
Qy	148	GGAGAGAGATTCTGCTCAATCGCGTGGCGGCAAGTCTATGCGATCCAGGACAGTGCC	207	
Db	496	GGGACAAGGTACTGCGTATGCGCGACGGGTCCTTCTCAAGCTGATTGCGCCGAAGCGC	437	
Qy	208	CTGCATCGCGGTGTAACGGCTTTCGAGCGGGTCGAGTGCCTATTCCAAAGAACCACTATCC	267	
Db	436	CTGGTTTCGCGCGAGTCCCTGTTGCCCTATCAGAAGCGCTTCGACTGCATGCCAGCGCG	377	
Qy	268	TGTTGGTATCAACGGCTCGACATATCGGTGGGACGATCGCCGCTCGTGGATATCTCACA	327	
Db	376	CTGCGCGGCGCGACATGGTCAGCGCCAAACGTGGTGGCCCTCGTCTCGCCGACAA	317	
Qy	328	AACCCCGGCGATGTCAGATCGGCGGCGCGCTTTGAGAGCTTCCCGGTGAAGAGGCC	387	
Db	316	CAGCGCTGCACGGTGGCATTTGACACAGCTCCGCTGTAGAACATGCACGTCGTCAAGCCAC	257	
Qy	388	AAAGGTCCTTATCTTCGTTTACGTAGGGGACGGCGAAACCAACCGCGCTTATCGAAGATGTG	447	
Db	256	ACGNACCGGAGTAGCTCAACGCACTCCATGCGCTCCAGCCGAGCTGATCGCCGCGATC	197	
Qy	448	CCGCGCGGCTTCCTTGATG	466	
Db	196	CAGCAACAGACGCTGTACG	178	

RESULT 4

LOCUS	BH872653	493 bp	DNA	linear	GSS 05-AUG-2002
DEFINITION	hp07e02.b2 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays genomic clone hp07e02 5', genomic survey sequence.				

ACCESSION BH872653
 VERSION BH872653.1 GI:22108550
 KEYWORDS GSS.

SOURCE
ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 493)
 REFERENCE

REFERENCE AUTHORS

TITLE Genomic shotgun sequences from Zea mays (methyl-filtered)
JOURNAL Unpublished (2002)

COMMENT

CONTACT: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA

tel: 516 367 8884
Fax: 516 367 8874
Email: mcconbie@cshl.org
Plate: hp07 row: e column: 02
Seq primer: -21M13UnivFwd
Class: shotgun
High quality sequence stop: 493

FEATURES

```

1. : 435
   /organism="Zea mays"
   /mol.title="genomic DNA"
   /cultivar="B73"
   /db_xref="taxon:4577"
   /clones="hp07e02"
   /lab_host="JMI07 or DH5
   /clone.lib="WGS-ZmaysF
   /notes="Organ: immature
   The vector was digested
   added by fill in in the
   was nebulized, and rep
   fractionated using seph
   between 0.8 and 3 kb an
   (x/y reads in M13mp19,
   ligation was transforme

```

ORIGIN

Query Match	3.6%	Score 40.4	DB 28	Length 493
Best Local Similarity	51.1%	Pred. No. 60		
Matches 95	Conservative 0	Mismatches 91	Indels 0	Gaps 0
Qy	372	CCCGGTTCAAAGCGCCAAAGTGCTTATCTTCGGTTTACGTAGGCGAGCGGCGAACCAACGCC	431	
Db	203	CCGGGATGTCCGGCACCAAGCGCTTCGGCGGTTCGGCCGTGGCGACCAAGCGCAGCGCC	262	
Qy	432	GCTTATCGAAGATGTCCGCCCGCGCTTCCTTGATGAAACCGCGCCATTTCACGGGCAACA	491	
Db	263	CGATGTACTCCATGTCGCGCGCGCTCCGCGCTTGGTGACGACGGCCACAGAGCCCGCAGC	322	
Qy	492	TCGGGCTCGTGCCCTCGCAACTGGCGCTTGGGTGCGGAAACCGCTTTTGATGCGGGGCACGT	551	
Db	323	TCTCGTGTAGCGCGCCGACACACGCTGGTGTCGGAACTCTCCGGCTTTGATCTCGAACTCGG	382	
Qy	552	CTTCAT	557	
Db	383	CGGCGT	388	

RESULT 5
BH872654/c

LOCUS	BH872654	626 bp	DNA	linear	GSS 05-AUG-2002
DEFINITION	hp07e02.g2 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays genomic clone hp07e02 5' genomic survey sequence.				

ACCESSION
 VERSION
 KEYWORDS

SOURCE
ORGANISM

zeae may be
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
ciade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS Rabinowicz P.D., O'Shaughnessy A.L., Balija V., Dedhia N., Katzenburger, F., King L., Miller B., Muller S., Nascimento, L., Zutavern, T., McConbie, W.R. and Martienssen, R.A.
TITLE Genomic shotgun sequences from *Zea mays* (methyl-filtered)
JOURNAL Unpublished (2002)

COMMENT

Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org

Plate: hp07 row: e column: 02
 Seq primer: -12M13UnivFwd
 Class: shotgun
 High quality sequence stop: 626.
 Location/Qualifiers

FEATURES

1..626
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="hp07e02"
 /lab_host="JM107 or DH5a"
 /clone_lib="WGS-ZmaysF (CM107 adapted methyl filtered)"
 /note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
 The vector was digested with XbaI and one nucleotide was
 added by fill in in the recessive 3' end. The genomic DNA
 was nebulized, end repaired, adaptor ligated and size
 fractionated using sephadex. The resulting fragments were
 between 0.8 and 3 kb and were cloned into the vector
 (x/y reads in M13mp19, b/g reads in pUC19). The same
 ligation was transformed in either JM107 or DH5a."

ORIGIN

Query Match 3.6%; Score 40.4; DB 28; Length 626;
 Best Local Similarity 51.1%; Pred. No. 67;
 Matches 95; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 372 CCCGTTGAAGAGGCCAAAGGCTTATCTTGTACGTAGCGACGGCGAACCAAGCC 431
 |||||
 Db 533 CCGGATGTCGGGACCAAGGCTTATCTTGTACGTAGCGACGGCGAACCAAGCC 474
 |||||
 QY 432 GCTTATCGAAGATGTCGGCGCGGCTTCTTGATGAAACCGCGCATTCACGCGCAACA 491
 |||||
 Db 473 CGATGACTCCATGTCGGCGCGGCTTCTTGATGAAACCGCGCATTCACGCGCAACA 414
 |||||
 QY 492 TCGGCTCGTGCGCTCGAACTGCGGTCGGTGGCGGAAACCGCTTGTATCGGGGCAAGT 551
 |||||
 Db 413 TCTCGTGTAGCGCGCGACACGACGCGTGTGCGCAACTCGCGCTTGATCTCGAACTCGG 354
 |||||
 QY 552 CTTTCAT 557
 |||||
 Db 353 CGGCGT 348
 |||||

RESULT 6
 LOCUS CC731706 841 bp DNA linear GSS 23-JUN-2003
 DEFINITION OGUDQ66TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMEMa0416L11,
 genomic survey sequence.

ACCESSION CC731706
 VERSION CC731706.1 GI:32151029
 KEYWORDS GSS.
 SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE

JOURNAL Consortium for Maize Genomics
 COMMENT Unpublished (2002)
 Other_GSSs: OGUDQ66TV
 Contact: Cathy Whitelaw

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TF
 Class: sheared ends.

FEATURES

Location/Qualifiers

source

1..841
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMEMa0416L11"
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb"
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb"
 methylation filtered genomic DNA library"

ORIGIN

Query Match 3.6%; Score 40.4; DB 29; Length 841;
 Best Local Similarity 51.1%; Pred. No. 77;
 Matches 95; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 372 CCCGTTGAAGAGGCCAAAGGCTTATCTTGTACGTAGCGACGGCGAACCAAGCC 431
 |||||
 Db 615 CCGGATGTCGGGACCAAGGCTTATCTTGTACGTAGCGACGGCGAACCAAGCC 674
 |||||
 QY 432 GCTTATCGAAGATGTCGGCGCGGCTTCTTGATGAAACCGCGCATTCACGCGCAACA 491
 |||||
 Db 675 CGATGACTCCATGTCGGCGCGGCTTCTTGATGAAACCGCGCATTCACGCGCAACA 734
 |||||
 QY 492 TCGGCTCGTGCGCTCGAACTGCGGTCGGTGGCGGAAACCGCTTGTATCGGGGCAAGT 551
 |||||
 Db 735 TCTCGTGTAGCGCGCGACACGACGCGTGTGCGCAACTCGCGCTTGATCTCGAACTCGG 794
 |||||
 QY 552 CTTTCAT 557
 |||||
 Db 795 CGGCGT 800
 |||||

RESULT 7

LOCUS CC731699/c 869 bp DNA linear GSS 23-JUN-2003
 DEFINITION OGUDQ66TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMEMa0416L11,
 genomic survey sequence.

ACCESSION CC731699
 VERSION CC731699.1 GI:32151015
 KEYWORDS GSS.
 SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE

JOURNAL Consortium for Maize Genomics
 COMMENT Unpublished (2002)
 Other_GSSs: OGUDQ66TV
 Contact: Cathy Whitelaw

FEATURES

Location/Qualifiers

1..869
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMEMa0416L11"
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb"
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb"
 methylation filtered genomic DNA library"

FEATURES

Location/Qualifiers

ORIGIN

Query Match 3.6%; Score 40.4; DB 29; Length 869;

```

Best Local Similarity 51.1%; Pred. No. 78;
Matches 95; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 372 CCGGTTGAAGGCAAGGCTTATCTTCTGTTACGTAGCGACGGCGAACCC 431
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
816 CCGGATGTCGGGCAACACGGCTCCGGCTTCTCCCGCTGGCGACCGACGCC 757
Qy 432 GCTTATCGAAGATGTCGCCCGCGCTTCTTCATGAAGAACCGGCCATT 491
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
756 CGATGTACTCCATGTCGCCCGCGCTCCGCCCTTGGTGACGCGCCAGAGCCG 697
Qy 492 TCGCTCTGCGCTCGAATCGCGCTTGGTGGCGGAAGACGGCTTGTATGCGG 551
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
696 TCTCGTGTAGCGCGCGACAGCAGGTGGTCCGAATCTCCGGCTGTATCTCG 637
Qy 552 CTTTCA 557
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
636 CCGCGT 631

RESULT 8
BG281668 916 bp mRNA linear EST 21-FEB-2001
LOCUS 602402350F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4544581 5',
DEFINITION mRNA sequence.
ACCESSION BG281668
VERSION BG281668.1 GI:13030594
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 916)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCW128 row: C column: 14
High quality sequence stop: 646.
FEATURES
Location/Qualifiers
1..916
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4544581"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 20"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected 500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN
Query Match 3.5%; Score 40; DB 12; Length 916;
Best Local Similarity 44.0%; Pred. No. 18+02;
Matches 166; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

Qy 161 TGTCAATCGCTGGCGCGCAAGGCTATGCGATCCAGACAGGTCCTCGCATCGCGTG 220
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
228 TGGTAGTCTCCAGCATCTGTGCCAAGTGGCCCGGACCGAGCCCGTGAAGTAATGCG 287

```

```

Qy 221 TAACGCTTCCGACCGGTCGAGTGTATTTCCAGACACCATATCTCTGCTGATCAG 280
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
288 AGTTCGTGGAGAACTGCAGACTTGGATACTGATTTATGGCTCAGGCTACCCCAATGATC 347
Qy 281 GCTGGACATATCGCTGGACGATGGCGCTCGTGCATATCTCACAACCCCGGCGAGTG 340
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 CCAAGACAAAGGCTGTGTTGAAGGACACGTGAGGCTGTGTTCGGCTTCCCGCAGTTTG 407
Qy 341 TGCAGATCGCGCGCGCTTTGAAGACGTTCCGGTTGAAGAGGCCAAGGCTTATCT 400
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 TCCGGTTCACTGGCGCACGNCAGACCATCTCGGAAAGAGGCGGAAGATGTTATAT 467
Qy 401 TCGTTTACGTAGCGACGCGGAACCAACGCGCTTATCGAAGATGTCCCGCGGCTTCC 460
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
468 GGGAGGACTCAGCATCCGAGAAATCAGAGGAGACTCAGGAAGATCACATCTTCTCTCA 527
Qy 461 TTGATGAAGAACCGCGCATTCAGGCCAACATCGGCTCGTGGCTGAACTGGCGTTGG 520
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
528 ATGAAGGTCCTCCAGCCCGCTCCCGCTTCTCCAGCGATATTTCTGGAACGCGGCTTGA 587
Qy 521 GTGCGGAAACCGGCTTT 537
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
588 GTCAGCAACCAAGCTCT 604

RESULT 9
BX424977/c
LOCUS BX424977 Homo sapiens PLACENTA Homo sapiens cDNA clone CLOBA005ZH04
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX424977
VERSION BX424977.1 GI:30784421
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 515)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
JOURNAL Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6304.f. For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CLOBA005ZH04FPI&cluster=6304.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CLOBA005ZH04FPI.
FEATURES
Location/Qualifiers
1..515
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CLOBA005ZH04"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
ORIGIN
Query Match 3.5%; Score 39.6; DB 13; Length 515;
Best Local Similarity 8.6%; Pred. No. 95;
Matches 16; Conservative 99; Mismatches 71; Indels 0; Gaps 0;

Qy 331 CCGGCACTGTGACATCGCGCGCGCTTTGAAGACGTTCCCGGTTGAGAGGCCA 390

```


derived from the human host cells."

ORIGIN
Query Match 3.4%; Score 39; DB 14; Length 547;
Best Local Similarity 46.2%; Pred. No. 1.4e+02;
Matches 129; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 648 GGTTCCTCGGCGCAAGCCCAAGAGTGTTCAGATCTGCTTGGCGAGCATTCGGTCCGGT 707
Db 16 GCTCGCGGGAACAAGATGCTGGCGCAAGTTCGACCTGGTGGCATCTCTCGGCTCCGAG 75
QY 708 TTTCGAAGGCATGATCGAAGGCAACCTGCAATCATCGCAACATTCGCGACGAGCGGT 767
Db 76 AGGCGCTTCGCGAGATCGAAGTTCATCTTCGACGTAGATCGAAGCGGTATCATGAACATCAG 135
QY 768 CGCCATCAGCATATCGATCTGCTCGCGGGTCTCTCAAGTTCGAAACCGTGCSCGATCC 827
Db 136 CCCCCTCGAAGACATCTGGCAAGCGACAGGAATACAGATTCAGTCGAGCGCGGCTT 195
QY 828 CGAGCTCAGCGAGTTCGAATGTAGTGCCTGCGGTGATGAGACCAACCTCTACTTCCA 887
Db 196 GTCGACTCACAAATGAGCAGATGTCAGGACGCGAGATGTACAAAGGACGAAGACA 255
QY 888 GACGCTGGCAAGTCGTGACGTCAAGAGGACGCGCAGA 926
Db 256 GAAGAAGAAGGACGCGCTGCAGGCGAAGAATGAGGCAGA 294

RESULT 12
LOCUS CA146289 697 bp mRNA linear EST 24-SEP-2003
DEFINITION SCVPR2079B04.9 RT2 Saccharum officinarum cDNA clone SCVPR2079B04
5', mRNA sequence.
ACCESSION CA146289
VERSION CA146289.1 GI:35045175
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum

REFERENCE
AUTHORS Vettores, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Contact: Arruda P
Centro de Biologia Molecular e Engenhariaia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bccccenter.fcav.unesp.br
Plate: 079 row: B column: 04
Seq primer: T7 Promoter Primer.

FEATURES
source
1..697
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCVPR2079B04"
/lab_host="DH10B"
/clone_lib="RT2"
/notes="Organ: Root tips (0.3cm-long) from adult plants;
Vector: pSport1; Site 1: SalI; Site 2: NotI; An
unidirectional cDNA library generated from [Root
tips (0.3cm-long) from adult plants]. cDNA was prepared
from polyA+ mRNA using SuperScript Plasmid System Kit
(Invitrogen). The double-strand cDNAs were fractionated
in a sepharose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 Kb were directionally cloned into the

vector. Details of each source of RNA and library
construction can be obtained at
http://sucet.lad.ic.unicamp.br/public"

ORIGIN
Query Match 3.4%; Score 39; DB 13; Length 697;
Best Local Similarity 45.2%; Pred. No. 1.5e+02;
Matches 141; Conservative 0; Mismatches 171; Indels 0; Gaps 0;
QY 576 GGTGAAGGGCAAGCATATCATCTTCGCGCTTGGCTTGGCCCTGGCGATCCGACAGCT 635
Db 279 GGAGAAGAGCGGCGAGATCACCTTCGGAGAAATGTCATCTACATCGTCGAGAACAATCAACGT 338
QY 636 TAGCGCTTCGAGGTTGCTGCGGGCAAGCCCAAGGTGTTTACGATCTGCTTGGCGAGCA 695
Db 339 CGCGGCATCGAACCAAGCTCTGTGCTCCATCGAGTGGCGCTCGCAGAGGTTGTCAACCA 358
QY 696 TTGCGTGCCTGGTTTTCGAAGGCATGATCGAAGCAACCTGCATCCATCGGCAACATTGG 755
Db 399 CCGCGCGGTGCAGCGCAAGGTCCGCGACGAGATCAAGGCGCTGCTCGGGGACGACGAGCC 458
QY 756 CAGCAAGCGCGTCCGATCATGATATCGATCTGCTGCGCGGCTACTCAAGTCCGAACC 815
Db 459 CATCAGAGTCCACCATCCAGAGCTCCCTACCTGCGAGCGGCTCATCAGGAGAGCT 518
QY 816 GTGCGCGGATCCCGAGCTCAGCAGTTCGAATGTGTACGTGCCGCTCGATGAGACGAGCA 875
Db 519 GCGCTCCACTCCCATCCGCTCTCTGCTCCGCGACATGAACCTCGAGAGGCCAAGCT 578
QY 876 CCTCTACTTCCA 887
Db 579 CGCGCGGCTACNA 590

RESULT 13
LOCUS CB028106 761 bp mRNA linear EST 13-JAN-2003
DEFINITION TgESTzyd47h02.y1 TgRH Tachyzoite Norm 7 cDNA Library toxoplasma
gondii cDNA clone TgESTzyd47h02.y1, similar to TR:Q24935 Q24935
ORGANELLAR HEAT SHOCK PROTEIN.1; mRNA sequence.
ACCESSION CB028106
VERSION CB028106.1 GI:27724478
KEYWORDS EST.
SOURCE Toxoplasma gondii
ORGANISM Toxoplasma gondii

REFERENCE
AUTHORS Tang, K., Cole, R., Fogarty, S., Sibley, L.D., Ajioke, J.A., White, M.,
Clifton, S., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M.,
Hallier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M.,
Ritter, E., Bennett, J., Franklin, C., Tsagareishvili, R., Ronko, I.,
Kennedy, S., Maguire, L., Waterston, R. and Wilson, R.
TITLE Toxoplasma EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Clifton, S.
Toxoplasma EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
Contact David Sibley (toxoe@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 441.

FEATURES
source
1..761
/organism="Toxoplasma gondii"
/mol_type="mRNA"
/strain="RH (Type 1)"
/db_xref="taxon:5811"
/clone="TgESTzyd47h02.y1"

/dev stage="Tachyzoite"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/clone_lib="TGRH tachyzoite Norm 7 CDNA Library"
/note="vector: paluscript SK; Site 1: EcoRI; Site 2:
xhoI; Toxoplasma RH strain tachyzoites were grown in human
foreskin fibroblast cultures in vitro. The library was
originally constructed by K.L.Wan, Cambridge University.
cDNAs were synthesized from polyA RNAs by oligo d(T)
priming and directionally cloned into the EcoRI to XhoI
sites of the lambda ZapII vector using the ZAP-cDNA
synthesis kit (Stratagene). The primary CDNA library was
mass excised as phagemid using EXAssist helper phage
(Stratagene). Phagemid DNA was extracted by
phenol-chloroform method, and hybridized against a pool of
highly abundant genes which were derived from short-cycle
PCR of the primary CDNA library. The normalized library
was electroporated into DH10B (GeneHog, Invitrogen, Inc).
WARNING: the library contains a small percentage of cDNAs
derived from the human host cells."

ORIGIN

Query Match 3.4%; Score 39; DB 14; Length 761;
Best Local Similarity 46.2%; Pred. No. 1.6e+02;
Matches 129; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 648 GGTGCTCGGGCAAGCCCAAGTGTTCAGATCTGCTTGGCGAGCATTCGGTGGCGGT 707
Db 168 GGTGCTCGGGCAAGCCCAAGTGTTCAGATCTGCTTGGCGAGCATTCGGTGGCGGT 227
QY 708 TTTCGAGGCATGATCGAAGCAACCTGCATCATCGCAACATTCGCGAGCGCGT 767
Db 228 AGGCTTCGCGAGATCGAAGTTCATTCGAGTGTAGTGCAGACGGTATCATGAACATCAG 287
QY 768 GCCATCAGCATATGATCTGGCTGCGGGCGTACTCAAGTGCAGACCGTGGCGGATCC 827
Db 288 GCGCTGACAGAGCACTGGCAAGCAGACAGAAATCAGATTGATGCGAGCGCGCTT 347
QY 828 CGAGCTCAGGATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 887
Db 348 GTCCGACTCACAATGTAGAGATGTGTCAAGAGCGCGGATGTACAGGAGCAGACGA 407
QY 888 GACGTGCGGCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 926
Db 408 GAAGAAGAAGAGCGCGGTGCGAGCGAGCAAGATGAGGCGAGA 446

RESULT 14

CNS01213/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BAC080C07 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 645)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaut at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector

pBelobAC11
Location/Qualifiers
1. .645
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BAC080C07"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : T7"

ORIGIN

Query Match 3.4%; Score 38.8; DB 29; Length 645;
Best Local Similarity 12.7%; Pred. No. 1.6e+02;
Matches 65; Conservative 125; Mismatches 322; Indels 1; Gaps 1;
QY 158 TTCTGCTCAATCGCTGGCGGCAAGTCTATGCATCCAGACAGAGTGCCTGCTATCGG 217
Db 604 TTNTTTTTTTSSSSSSSSCAASSSSNCAGNNSSCNGNNSSNCSSSVACAGSS 545
QY 218 GTGTAACGCTTTCCGACCGGTCGAGTGTCTATTCGAAGAACCACCATATCTCTGCTATC 277
Db 544 SSANCCSSACSCSNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSS 485
QY 278 AGGCTGGACATATCGCTGGAGCAGTGGCGGCTCGTGCATATCTCAACAACCCCGCA 337
Db 484 MNSSAAANSSNAASASNTTNSSSNNNNNNSSNNSSNNSSNNSSNNSSNNSSNNSS 425
QY 338 GTGTGAGATCGCGCGCGGCTTTGAAGACGCTTCCGCTTGAAGAGGCAAGGCTCTTA 397
Db 424 SNAASSSSAASNS 365
QY 398 TTTTC-GTTTACGTAGCGAGCGGCAACCAACCGCGCTTATCGAAGATGTGCGCGCGGC 456
Db 364 VSNNTNNNNNNSSSSSSSSSVACSSSSSSSCATSSNSNSNSNSNSNSNSNSNSNS 305
QY 457 TTCCTTGTGAAAAACCGCGCCATTCAAGGCCAATCGCTCGCTCGAATCGGCGC 516
Db 304 NNSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNN 245
QY 517 TTGGGTGCGAAACCGGCTTTCATCGGGGCGACGCTTCATTCACAAGATTCGATCTGT 576
Db 244 SSCSSNNNNNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSS 185
QY 577 GTGAGGGCAACGATATCATCTCGCGCTTGGCTTGGCGCTGGCGGATCCCGACGCTT 636
Db 184 CNNAAACCNNAANNN 125
QY 637 ACAGCTTCGAGGTTGCTCGCGGCAAGCCCAA 669
Db 124 ANNNGGCCNN 92

RESULT 15

AJ489090/c
LOCUS
DEFINITION
AJ489090 Citrus clementina x Citrus reticulata flavedo Citrus
clementina x Citrus reticulata cDNA clone ACR833, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Citrus clementina x Citrus reticulata
Citrus clementina x Citrus reticulata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 657)
Sanchez-Ballesta M.
Gene expression and heat-induced chilling tolerance in Fortune
mandarin fruit
Unpublished (2003)
Contact: Sanchez-Ballesta M
Ciencia Alimentos-Postcosecha

Qy 914 AGGAAGC 920
|||
Db 3253973 TGGGTGC 3253979

RESULT 2

US-09-103-840A-1
; Sequence 1, Application US/09103840A

; Patent No. 6294328
; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; FILE REFERENCE: 107196.136

; CURRENT FILING DATE: 1998-02-18

; PRIOR FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 4411529

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; OTHER INFORMATION: H37RV

US-09-103-840A-1

Query Match 4.1%; Score 46.2; DB 3; Length 4411529;
Best Local Similarity 46.9%; Pred. No. 0.036;
Matches 144; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Qy 614 CGCTGGCGATCCCGACGAGCTTACGGGTTCCGAGGTTGCTGGGGCAAGCCCAAGGTG 673

Db 3259397 CGCTGGAAGACCGCGGAGCTACGGGTGCCCTGTTGATCGGTGACCCCGCGCG 3259456

Qy 674 TTACGATCTGCTGGGAGCATTCGGTGGCGGTTTTCGAAGGATGATCGAAGGCAAC 733

Db 3259457 GTGACCGGCTGCTCCCGCGGTGCGATGCTGCTGCGCGACCGCATCACGAGCTCGAGC 3259516

Qy 734 CTGCAATCCATGGCAACATTGGCAGCAAGCGGTGCGCATCAGCATATGATCTGCTGC 753

Db 3259517 TGGCATCGCGCGGAGCAAGCAAGCTGCGCGGCGATCAGCCGACCTCTGGAGC 3259576

Qy 794 CGGCGCTACTAAGTTCGAACCGTGGCGCGATCCGAGCTCAGCATTCGAATGTAGC 853

Db 3259577 GGATCGTTGTGTTGTGTCGCGCGCGAGCAACGAAATCGATCCCGAGAGGCTCAAC 3259636

Qy 854 TGCGGTCGATGACACGACGACCTTACTTCCAGACGCTGGGCAAGTGTGACGTCRA 913

Db 3259637 TGGAGTTGGACGACGACGACGCTGCTGTCGCCAGCTGTTGAGACCGTACCGGAA 3259696

Qy 914 AGGAAGC 920

|||

Db 3259697 TGGGTGC 3259703

RESULT 3

US-09-252-991A-12390/c

; Sequence 12390, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; NUMBER OF SEQ ID NOS: 33142

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 3024

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12390

Query Match 3.5%; Score 40; DB 4; Length 3024;
Best Local Similarity 47.0%; Pred. No. 0.088;
Matches 124; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 643 TCCGAGGTTGCTGGCGGCAAGCCCAAGGTTTACGATCTGCTTGGGAGCATTCGGTG 702

Db 814 TCGCGGTGGCGGTGGAGTCGATCCATCGCGTTGGGACGCGCTGCCGGCAATACCGC 873

Qy 703 CCGGTTTTCGAAGGATGATCGAAGGCAACCTGCAATCCATGGCAATTTGGCAGCAAG 762

Db 874 GAGTGGCCCCAGGACGATCGGATGGCTGAAGGTCACCTGGATGCCCTGCGGCCAG 933

Qy 763 CCGCTGGCATCAGCATATCGATTCGGTGGCGGCGTACTCAAGTTCGAACCGTGGCG 822

Db 934 GGCATGCCCATCGTCGCCATCGACTTACCTGCGCGGAGCGCGGAGCGCGCGCG 993

Qy 823 GATCCGAGCTCACGACGTTGCAATGGTACGTCGCGGTGATGAGACGACCACTCTAC 882

Db 994 CTCGCTGGCGCTGCTGATGAGGATCGTCCGCTTCGTCAGCACCCCGCGCTGGAC 1053

; SEQ ID NO 12390

; LENGTH: 2493

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12390

Query Match

Best Local Similarity 3.5%; Score 40; DB 4; Length 2493;

Matches 124; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 643 TCCGAGGTTGCTGGCGGCAAGCCCAAGGTTTACGATCTGCTTGGGAGCATTCGGTG 702

Db 2355 TCGCGGTGGCGGTGGAGTCGATCCATCGCGTTGGGACGCGCTGCCGGCAATACCGC 2296

Qy 703 CCGGTTTTCGAAGGATGATCGAAGGCAACCTGCAATCCATGGCAATTTGGCAGCAAG 762

Db 2295 GAGTGGCCCCAGGACGATCGGATTTGGCTGAAGGTCACCTGGATGCCCTGCGGCCAG 2236

Qy 763 CCGCTGGCATCAGCATATCGATTCGGTGGCGGCGTACTCAAGTTCGAACCGTGGCG 822

Db 2235 GGCATGCCCATCGTCGCCATCGACTTACCTGCGCGGAGCGCGGAGCGCGCGCG 2176

Qy 823 GATCCGAGCTCACGACGTTTCAATGGTACGTCGCGGTGATGAGACGACCACTCTAC 882

Db 2175 CTCGCTGGCGCTGCTGATGAGGATCGTACGTCGCGTTTCGTCAGCACCCCGCGCTGGAC 2116

Qy 883 TTCCAGACGCTGGGCAAGTCTGTG 906

Db 2115 TACCTGGGGTGAGCGACGTCGAG 2092

RESULT 4

US-09-252-991A-12202

; Sequence 12202, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; NUMBER OF SEQ ID NOS: 33142

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 12202

; LENGTH: 3024

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12202

QY 883 TTCCAGACGCTGGGCAAAAGTCGTG 906
DB 1054 TACCTGGGGTGAGCGACGTCGAG 1077

RESULT 5

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 3.4%; Score 39.2; DB 3; Length 4403765;
Best Local Similarity 45.7%; Pred. No. 4.2; Indels 0; Gaps 0;
Matches 137; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
QY 641 GTTCCGAGGTTGCTGGGGCAAGGCTTTTACGATCTGCTTGGCGAGCATTCGG 700
DB 3139910 GTACCGTCGAGGCGCTGGAAGGCCCTGATGGTATCCAGTGGAGCAGAGTGGTG 3139851
QY 701 TGGCGGTTTTCGAGGATGATCGAAGGCAACCTGCAATCCATGCGCAACATTTGGCAGCA 760
DB 3139850 TGGCGGTGATCGACCGCGGTGCGCGGATTAACCGAACCACGTCACACCTGGCGTCGG 3139791
QY 761 AGCGCTCGCCATCAGCATATCGATCTGGCTGCGGGCGTACTCAAGGTGCAACCGTGGC 820
DB 3139790 CTTCCGATGGGTGATCATCGGTTTCAATGTGGCGCCGAAAGCGACCGAGCTGG 3139731
QY 821 CGGATCCGAGCTACGCGAGTTGCAATGTGATGTCGCGGTGCGATGAGACCGACCTCT 880
DB 3139730 CCAGCCGGAAGCGGTGGAGATCCGCTACTCTCGGTCTATCACCAGGCGATCGACGAGA 3139671
QY 881 ACTTCCAGACGCTGGGCAAAAGTCGTGACGTCAAAGGAAGCGCAGACTCTTTCGAGCGAG 940
DB 3139670 TCGAGCAGGCGCTGCTGGCTGCTCAAGCCGATCTACAGGAAACAGCTGGGTGGG 3139611

RESULT 6

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1

; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 3.4%; Score 39.2; DB 3; Length 4411529;
Best Local Similarity 45.7%; Pred. No. 4.2;
Matches 137; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 641 GTTCCGAGGTTGCTGGGGCAAGGCTTTTACGATCTGCTTGGCGAGCATTCGG 700
DB 3145742 GTACCGTCGAGGCGCTGGAAGGCCCTGATGGGTATCCAGTGGAGCAGAGTGGTGC 3145683
QY 701 TGGCGGTTTTCGAAGGATGATCAAGGCAAACTGCAATCCATGGAATGGCAATTTGGCAGCA 760
DB 3145682 TGGCGGTGATCGACCGCGGTGCGCGGCATTAACGAAACCAACGTCACCTGGCGTCGG 3145623
QY 761 AGCGCTCGCCATCAGCATATCGATCTGGCTGCGGGCGTACTCAAGGTGCAACCGTGGC 820
DB 3145622 CTTCCGATGGGTGATCATCGGTTTCAATGTGGCGCCGGAAGCGACGACGAGCTGG 3145563
QY 821 CGGATCCGAGCTCACGCGAGTTGCAATGTGACGTGCGGTGCGATGAGACCGACCTCT 880
DB 3145562 CCAGCCGGAAGCGGTGGAGATCCGCTACTCTCGGTCTATCACCAGGCGATCGACGAGA 3145503
QY 881 ACTTCCAGACGCTGGGCAAAAGTCGTGACGTCAAAGGAAGCGGCGAGACTCTTTCGAGCGAG 940
DB 3145502 TCGAGCAGGCGCTGCTGGGCTGCTCAAGCGGATCTACGAGGAAACACGCTGGGTGGG 3145443

RESULT 7

US-09-489-039A-2455
; Sequence 2455, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2455
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2455

Query Match 3.4%; Score 38.4; DB 4; Length 1053;
Best Local Similarity 51.2%; Pred. No. 0.16;
Matches 171; Conservative 0; Mismatches 151; Indels 12; Gaps 3;

QY 85 TGGTATCCCGTTCGCCCTCGCGAGCGAAATCGCCGAAGGTACTCCCGTCCCGTCAAGCTC 144
DB 91 TGGCACCCTGGTGGCGCGCTGCGATATTGCC---GGCGCCCGCGGTGAAAGCCACCCCTG 147
QY 145 CTGGGAGAGAGATTCTGCTCAATCGCTGGGGGCGAAGGTCTATGCGATCCAGGACAGG 204
DB 148 CTGGACGAACAGCTGGTTATCTATCGTATTAAAGGCGAGGTGGTGGTCCCGCGGACGTC 207
QY 205 TGCCTGCATCGCGGTGTAAAGCGCTTTCCGACCGGCTCGAGTGCATTTCCAAAGAACACCAT 264
DB 208 TGCCCGCACCAGCGGGGTGCGCTGACCC---TGGGATTTTCATGAAGAGGAGGCATC 261
QY 265 TCCTGTGTTGATCAGCGCTCGACATATCGCTGGGACGATGCGCGCTCGTGCATATCTC 324
DB 262 GTCTGCCCTTATCATGCGCTGCGCTTTGGC---GAGGATGCGCGCTGCAACCGCATCCCC 318
QY 325 ACAAACCCGCGAGTGTGCAGATCGCGCGCGCGCTTTGAAGACGTTTCCCGGTTGAAGAG 384

Db 319 TCAGTCCGGGCAACCCATTCGGCCAAACTGCATCTCACCAGCTTCGGCGTGGAGGAG 378
Qy 385 GCCAAAGCTTATCTTCGTTTACGTAGGCGAGC 418
Db 379 CGTACGGGCTGATCTGGACCTGCTGCGCTGGC 412

RESULT 8

US-08-246-489-1/c
; Sequence 1, Application US/08246489
; Patent No. 6225049
; GENERAL INFORMATION:
; APPLICANT: Lan, Michael S.
; TITLE OF INVENTION: NOVEL HUMAN INSULINOMA-ASSOCIATED CDNA
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive
; CITY: Newport Beach
; STATE: California
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/246,489
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,715
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH012.012A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2838 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: IA-1

US-08-246-489-1

Query Match 3.3%; Score 37.4; DB 3; Length 2838;
Best Local Similarity 51.5%; Pred. No. 0.5;
Matches 86; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
Qy 267 CTGCTGTATCACGGCTGGACATATCTCGGACGATGGCGGCTCGTCGATATCTCTAC 326
Db 724 CGGCAGGGGCGAGTGGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 665
Qy 327 AACCCCGGAGTGTGAGATCGGCGGCGGGCTTTGAAGACGTTCCTCGGTTGAAGAGGC 386
Db 664 ACGCGTGCCCATCTTGAAGTTCGGCGGGGCGGAAGACGACGGGTGCGCCCGCGAGTGC 605
Qy 387 CAAGGCTTATCTTCTGTTTACGAGGACGGGACCAACCGCGC 433
Db 604 CGCTCCGCCAGTCCGCTCGGCGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 558

RESULT 9

US-09-620-312D-481
; Sequence 481, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Weinman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yuning
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: NO. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 481
; LENGTH: 6765
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(3105)
US-09-620-312D-481

Query Match 3.3%; Score 37.2; DB 4; Length 6765;
Best Local Similarity 56.6%; Pred. No. 0.85;
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
Qy 661 AAGCCCAAGGTGTTTACGATCTGTTGGCGAGCATTCGGTGGCGGTTTTCGAAGGCATG 720
Db 892 AGGCACCTAGCATCTCCAATATGCCATGTGTAATCTGTGACTGTGCTGAAGGACTC 951
Qy 721 ATCGAAGGCAACCTGCAATTCATGGCAACATTGGCGAAGCGCGTCGCCATCAGCAT 780
Db 952 ATAGAAGACCATTTTGTATGTCACCTGTAAGATGAGCACCTATCTGCTGGCTTCATCAT 1011
Qy 781 TC 782
Db 1012 TC 1013

RESULT 10

US-09-452-239-1/c
; Sequence 1, Application US/09452239
; Patent No. 6465229
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Fader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase
; FILE REFERENCE: B01284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1999-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50

```
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (806)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (810)
US-09-452-239-1

Query Match      3.3%; Score 37; DB 4; Length 891;
Best Local Similarity 50.9%; Pred. No. 0.39;
Matches 88; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 363 GAAGACGTTCCGGTTGAAGAGGCAAGGTTTATCTTCTTACGTAGCGAGCGGGA 422
DB 648 GGAGCGGCTCGGTAGCTGAGGTAGTTGTCTTGTGCGCGTCCACGAAGCGAAGTCA 589
QY 423 ACCAACGCGCTTATCGAAGATGTGCGCGCGCTTCTTGATGAAACCGCGCATCA 482
DB 588 AGACCGGCTGCTCTTGTCCGCCACGAGTGTCTCAGACGCGGAGCGGGGCCCT 529
QY 483 CGGCCAACATCGCTCGTGGCTCGAATGCGGCTTGGGTGCGGAAACGGCT 535
DB 528 CGCGAAGTCGATCTTGTGGCCACGCGCGCTTGTGATGAGGAAGGCT 476

RESULT 11
US-09-252-991A-14103
; Sequence 14103, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14103
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14103

Query Match      3.2%; Score 36.4; DB 4; Length 1869;
Best Local Similarity 45.9%; Pred. No. 0.82;
Matches 124; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 171 CGTGGCGGCAAGGTTCTATCGATCCAGACAGGTGCTGCATCGCGGTGTACGCTTC 230
DB 1720 CGCGGGCTGCTCGAGGATGTTCTCGAGCGGCTGCGCGCGGGTTTCCCGATCC 1661
QY 231 CGACCGGGTCGAGTGTCTATTCCAAAGAACACCATATCTGTGTATACAGGCTGGACATA 290
DB 1660 CGCGTGTATGTCGCTACGCGAGGCGGAGCAGCAGCCGAACTCGAATCGCCTGCGCTG 1601
QY 291 TCCTGGGAGATGCGCGCTGCTCATATCTTCAAAACCCCGGAGTGTGAGATCGG 350
DB 1600 CGCGCGCATCTCGAGTACTGTGTGAAACCTGGCGCGGAGTGGGATGTCCAGGCCA 1541
QY 351 CGCGCGGCTTTGAAGACGTTCCCGGTTGAAGAGGCAAGGTTCTTATCTTCTTTACGT 410
DB 1540 TCTGGCGGAGCTGGAGTGGCGGTTGTTGTGTGAGCGCGGCTGTCTGCGGAGTGTGCAACT 1481
QY 411 AGCGAGCGGCAACCAACGCGCTTATCGA 440
DB 1480 GGGCAGCGCGCGCAGCGCTGCGGACGA 1451

RESULT 13
US-09-380-420C-1
; Sequence 1, Application US/09380420C
; Patent No. 6300544
; GENERAL INFORMATION:
; APPLICANT: Halkier, Barbara
; Bak, Soren
; Kahn, Rachel
; Moller, Birger
; TITLE OF INVENTION: Cytochrome P450 Monooxygenases
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESS: Syngenta Patent Dept.
; STREET: 3054 Cornwallis Road
; CITY: RTP
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/380,420C

FILING DATE: 12-Jul-2001

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: S-21251A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1929 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

IMMEDIATE SOURCE:

CLONE: P450ox

FEATURE:

NAME/KEY: CDS

LOCATION: 81..1673

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-380-420C-1

Query Match 3.2%; Score 36.4; DB 4; Length 1929;

Best Local Similarity 45.5%; Pred. No. 0.83;

Matches 130; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 97 CGCTCGGAGCGAAATCGCGAAGTACTCCCGTCCCGTCAAGTCTCTGGGAGAGAG 156

Db 753 CGCTTCAGCACGTCGTGGACGCGCATGGACATGATGCGCAGCTTCTCCGCCGAGGAC 812

QY 157 ATTCTGCTCAATCGCTGGCGGCAAGTCTATGCGATCCAGGACAGGTGCTGCAATCGC 216

Db 813 TTCTTCCCAACGCGCGCGCGCTCGCGACGCGCTCTCGGGCTTCTCGCCCGCGC 872

QY 217 GGTGTAAAGCTTTCCGACCGGGTCGAGTGCTATTCCAAAGAACACCATATCTCTGTGAT 276

Db 873 GAGCGCATCTTCAACGAGCTCGACGCTTCTTCGAGAAGGTTCATCGACCAGCACATGGAC 932

QY 277 CAGCGCTGGACATATCGCTGGGACGATGGCGCGCTGTCGATATCTCTCAAAACCCCGC 336

Db 933 CCG 992

QY 337 AGTGTGAGATCGCGCGCGCGCGCGCTTTGAAGACGCTTCCCGTTGAAG 382

Db 993 AAGGAGCAGCAGCGCGCGCGCTTCCCGTTGAAGACGCTTCCCGTTGAAG 1038

RESULT 14

US-09-380-420C-1

Sequence 1, Application US/09899642A

Patent No. 6649814

GENERAL INFORMATION:

APPLICANT: Halkier, Barbara

Kahn, Rachel

Moller, Birger

TITLE OF INVENTION: Cytochrome P450 Monooxygenases

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: Syngenta Patent Dept.

STREET: 3054 Cornwallis Road

CITY: RTP

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/899,642A

FILING DATE: 05-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/380,420

FILING DATE: 12-Jul-1999

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: S-21251A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1929 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

IMMEDIATE SOURCE:

CLONE: P450ox

FEATURE:

NAME/KEY: CDS

LOCATION: 81..1673

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-899-642A-1

Query Match 3.2%; Score 36.4; DB 4; Length 1929;

Best Local Similarity 45.5%; Pred. No. 0.83; 156; Indels 0; Gaps 0;

Matches 130; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 97 CGCTCGGAGCGAAATCGCGAAGTACTCCCGTCCCGTCAAGTCTCTGGGAGAGAG 156

Db 753 CGCTTCAGCACGTCGTGGACGCGCATGGACATGATGCGCAGCTTCTCCGCCGAGGAC 812

QY 157 ATTCTGCTCAATCGCTGGCGGCAAGTCTATGCGATCCAGGACAGGTGCTGCAATCGC 216

Db 813 TTCTTCCCAACGCGCGCGCGCTCGCGACGCGCTCTCGGGCTTCTCGCCCGCGC 872

QY 217 GGTGTAAAGCTTTCCGACCGGGTCGAGTGCTATTCCAAAGAACACCATATCTCTGTGAT 276

Db 873 GAGCGCATCTTCAACGAGCTCGACGCTTCTTCGAGAAGGTTCATCGACCAGCACATGGAC 932

QY 277 CAGCGCTGGACATATCGCTGGGACGATGGCGCGCTGTCGATATCTCTCAAAACCCCGC 336

Db 933 CCG 992

QY 337 AGTGTGAGATCGCGCGCGCGCTTTGAAGACGCTTCCCGTTGAAG 382

Db 993 AAGGAGCAGCAGCGCGCGCTTCCCGTTGAAGACGCTTCCCGTTGAAG 1038

RESULT 15

US-09-252-991A-16243

Sequence 16243, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 16243

LENGTH: 1314

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16243

Query Match 3.2%; Score 36.2; DB 4; Length 1314;
Best Local Similarity 49.2%; Pred. No. 0.8;
Matches 95; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY	745	GGCAACATTGGCAGCAAGCGCTGCGCATCAGCATATCGATCTGGTGC	CGGGCGTACTC	804
DB	812	GGCGAGATCTTACCAAGGTCGACGACAGCTATATGGGCGTGACCTC	AGCGCATGATG	871
QY	805	AAGGTGAAACCGTGGCGCGATCCGAGCTCACCGAGTTCGAATG	TACGTGCCGTCGAT	864
DB	872	CTGCGCGCGCCAGCGCGCTGGCCGAGCGCTGGAGCTGACCGCT	TGGTGACCGCGAG	931
QY	865	GAGACCAAGCCACTCTACTTCCAGACGCTGGGCAAGTGTGACGT	CAAGGAAGGGCA	924
DB	932	GCGATCTCCAGGTGTCCAGCCAGACCTTCCGACCTCTCGGTGAT	CGACCGGGTTACC	991
QY	925	GACTCCTTCGAGC		937
DB	992	GACACCTGTGC		1004

Search completed: June 19, 2004, 07:22:48
Job time : 142.708 secs

QY 181 AAGTCTATGATCCAGGACAGGTGCTGCATCGCGGTGTAAAGCTTTCGACCGGGTC 240
Db |||||
QY 181 AAGTCTATGATCCAGGACAGGTGCTGCATCGCGGTGTAAAGCTTTCGACCGGGTC 240
Db |||||
QY 241 GAGTCTATCCAGGACACCATATCTGCTGTGTATCAAGGCTGGACATATCTGGAC 300
Db |||||
QY 241 GAGTCTATCCAGGACACCATATCTGCTGTGTATCAAGGCTGGACATATCTGGAC 300
Db |||||
QY 301 GATGGCGCTCTGTCGATATCTTCAAAAACCCCGGAGTGTGAGATGCGCGCGCT 360
Db |||||
QY 301 GATGGCGCTCTGTCGATATCTTCAAAAACCCCGGAGTGTGAGATGCGCGCGCT 360
Db |||||
QY 361 TTGAAGACGTTCCCGGTTGAAGAGCCAAAGCTCTATCTTCTGTTTACGTAGCGACGGC 420
Db |||||
QY 361 TTGAAGACGTTCCCGGTTGAAGAGCCAAAGCTCTATCTTCTGTTTACGTAGCGACGGC 420
Db |||||
QY 421 GAACCAACCGCGCTTATCGAAGATGTCCCGCGGCTTCTTGTATGAAAAACCGCGCAT 480
Db |||||
QY 421 GAACCAACCGCGCTTATCGAAGATGTCCCGCGGCTTCTTGTATGAAAAACCGCGCAT 480
Db |||||
QY 481 CACGGCCAAACATCGGCTCGTGGCTCGAACTGGCGCTTGGTGGGAAACCGCTTTGAT 540
Db |||||
QY 481 CACGGCCAAACATCGGCTCGTGGCTCGAACTGGCGCTTGGTGGGAAACCGCTTTGAT 540
Db |||||
QY 541 GCGGGGACGCTCTTCAATCAAGAAATTCGATCTCGTGAAGGCAACGATATCATCTCG 600
Db |||||
QY 541 GCGGGGACGCTCTTCAATCAAGAAATTCGATCTCGTGAAGGCAACGATATCATCTCG 600
Db |||||
QY 601 CCGTTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 660
Db |||||
QY 601 CCGTTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 660
Db |||||
QY 661 AAGCCCAAGGTTTACGATCTGCTTGGGAGCATTCGCTGCGGTTTTCGAAGGCAATG 720
Db |||||
QY 661 AAGCCCAAGGTTTACGATCTGCTTGGGAGCATTCGCTGCGGTTTTCGAAGGCAATG 720
Db |||||
QY 721 ATCGAAGGCAACCTGCAATCAATGCGAAATTTGGCAGCAAGCGCTGCGCATCAGCATA 780
Db |||||
QY 721 ATCGAAGGCAACCTGCAATCAATGCGAAATTTGGCAGCAAGCGCTGCGCATCAGCATA 780
Db |||||
QY 781 TCGATCTGGCTCGCGGCGTACTCAAGTTCGAACTGGCGGATCCGAGCTCAGCGAG 840
Db |||||
QY 781 TCGATCTGGCTCGCGGCGTACTCAAGTTCGAACTGGCGGATCCGAGCTCAGCGAG 840
Db |||||
QY 841 TTGAAATGATGTCGCGGTCGATGAGACAGGCACTTCTTCTTCAAGAGCTTGGGCAAA 900
Db |||||
QY 841 TTGAAATGATGTCGCGGTCGATGAGACAGGCACTTCTTCTTCAAGAGCTTGGGCAAA 900
Db |||||
QY 901 GTCGTGACGTCGAAAGGAGCGGAGACTCTTTCGAGCGAATTTCCACGAAATGCGGTA 960
Db |||||
QY 901 GTCGTGACGTCGAAAGGAGCGGAGACTCTTTCGAGCGAATTTCCACGAAATGCGGTA 960
Db |||||
QY 961 GGCCTCGGCTTAAACGCTTCAATGATGACGACATCATGCGACGTAATCGATGAGCGG 1020
Db |||||
QY 961 GGCCTCGGCTTAAACGCTTCAATGATGACGACATCATGCGACGTAATCGATGAGCGG 1020
Db |||||
QY 1021 TTCTACGCTGATGATCGGCTTGGTTCGAGAGAAATCTGTTTCGAGCGGAGCGCAATC 1080
Db |||||
QY 1021 TTCTACGCTGATGATCGGCTTGGTTCGAGAGAAATCTGTTTCGAGCGGAGCGCAATC 1080
Db |||||
QY 1081 ATCGAGTGGCGGGGCTTGGCAGTCAAGCAATCGCGGCTTTCAGGAGCAGCGTTGA 1137
Db |||||
QY 1081 ATCGAGTGGCGGGGCTTGGCAGTCAAGCAATCGCGGCTTTCAGGAGCAGCGTTGA 1137
Db |||||

RESULT 2

US-10-282-122A-28788
; Sequence 28788, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 28788
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28788

Query Match 4.1%; Score 46.2; DB 13; Length 3851;

Best Local Similarity 46.9%; Pred. No. 0.00029; Mismatches 163; Indels 0; Gaps 0;
Matches 144; Conservative 0;

QY 614 CQCCTGGCGATCCCGACAGCTTACGCGTTCGAGGTTGCTCGGCGCAAGCCCAAGGTG 673
Db |||||
QY 614 CQCCTGGCGATCCCGACAGCTTACGCGTTCGAGGTTGCTCGGCGCAAGCCCAAGGTG 673
Db |||||
QY 674 TTACGATCTGCTGGCGAGCATTCGCTGCGGTTTTCGAGGATGATCGAAGGCAAC 733
Db |||||
QY 674 TTACGATCTGCTGGCGAGCATTCGCTGCGGTTTTCGAGGATGATCGAAGGCAAC 733
Db |||||
QY 1069 GTGACCCGCTGCTGCCCGCTGAGTCTGCTGCGCGACCGCATCACCGACCTCGAGC 1128
Db |||||
QY 734 CTGCAATCCATGGCAACATTGGCAGCAAGCGCTGCGCATCAGCATATCGATCTGGGTGC 793
Db |||||
QY 1129 TGGCATCCGCGCGAGCAAGCCAGCTGCGCGGGGATCAGCGAACCTCTCTGGGACG 1188
Db |||||
QY 794 CGGCGGTAATCAAGTTCGAAACCGTGGCGGATCCGAGCTCAGCAGTTCGAAATGGTACG 853
Db |||||
QY 1189 GGATCGTTGCTGCTGCTGCCCGCCGAGCAACGAAATCGATCCGAGCAGGCTCAAC 1248
Db |||||
QY 854 TGCCGCTGATGAGACCAAGCCAGCTCTACTTCCAGACGCTGGCAAGTCTGACGCTCAA 913
Db |||||
QY 1249 TGGAGTGGCAGCACACGACGCTGCTGCTGCCAGCGTGTGCTGACGACCTGACCGAA 1308
Db |||||
QY 914 AGGAAGC 920
Db |||||
QY 1309 TGGGTGC 1315
Db |||||

RESULT 3

US-10-282-122A-26551

```
; Sequence 26551, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26551
; LENGTH: 5496
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-282-122A-26551

Query Match          4.1%; Score 46.2; DB 13; Length 5496;
Best Local Similarity 46.9%; Pred. No. 0.00033;
Matches 144; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 614 CGCCTGGCGATCCCGACCAAGTTACGCGTTCCGAGTTGCTCGCGGCAAGCCCAAGGTG 673
Db 2648 CCGTGGAAAAGACCGCGAGGTACGGTGCCCTGTTGATCGGTGACCCCGCGCGG 2707

QY 674 TTACGATCTGTTGGCGAGCATTCGGTCCGGTTTTGAAGCGATGATCGAAGCGCAAC 733
Db 2708 GTGACCGCTGTGCGCGGCTGCACTGCTCGTCGCGACCGCATCACGACCTCGAGC 2767

QY 734 CTGCAATCATCGACATTCGCAAGCGGTGCGCATCAGCATATCGATCTCGCTGC 793
Db 2768 TGGCATTCGCGCGCGACGAAGCCACGCTCGCGCGCGATCAGCGAAGCTCTCTGGAGC 2827

QY 794 CGGGCGTACTCAAGTTCGAACCGTGGCCCGATCCGAGCTCAGCGAGTTCGAATGTCG 853
Db 2828 GGATCGTTGTTCTGTCGCCGCCGAGCAACGACGATCGATCGCGAGCGAGGTCAAC 2887

QY 854 TGGCGTGCATGAGACCAAGCCACCTTACTTTCAGACGCTGGGCAAGTCTGTCGTCGA 913
Db 2888 TGAAGTTGACGACGACACGACCGCTGCTGCTGCCAGCGTGTGTCGACCGGTGACCGCAA 2947
QY 914 AGGAAGC 920
```

```
Db 2948 TGGGTGC 2954

RESULT 4
US-10-338-110-119
; Sequence 119, Application US/10338110
; Publication No. US20040023254A1
; GENERAL INFORMATION:
; APPLICANT: Fuhrmann, Jeffrey J.
; APPLICANT: Romesser, James A.
; TITLE OF INVENTION: A Method To Assess Quorum Sensing Potential Of Microbial
; TITLE OF INVENTION: Communities
; FILE REFERENCE: HER-0056
; CURRENT APPLICATION NUMBER: US/10/338,110
; CURRENT FILING DATE: 2003-01-07
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 119
; LENGTH: 536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus Sequence
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7)..(7)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (33)..(33)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (21)..(21)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (69)..(69)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (87)..(87)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (213)..(213)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (226)..(226)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (261)..(261)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (273)..(273)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (347)..(347)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (357)..(357)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (398)..(398)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (419)..(419)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (509)..(509)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (517)..(517)
; OTHER INFORMATION: n is a, c, g, or t
US-10-338-110-119

Query Match      3.8%; Score 43.6; DB 17; Length 536;
Best Local Similarity 13.2%; Pred. No. 0.00096;
Matches 67; Conservative 230; Mismatches 208; Indels 2; Gaps 1;

Qy 373 CCGGTTGAAGAGCCAAAGTCTTATCTTCGTTACGTAGGCGAGCGGCAACCAAGCCG 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 SVRHKTVMVMGVMVWVBTHNSYSRHKCTSGNSCRCTWYCGGCRYSRVSWVWVYTSRGR 60

Qy 433 CTTATCGAAGATGTGCGCGCGGCTTCCTTGATGAAAACCGCGCCATTACGCGCAACAT 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 MRRCVTGNTGSSVBYCVWYSKNVVDVMSBKWVGARYKSGAYCARTWCGAYWY 120

Qy 493 CGGCTCGTGCCTCGAACTCGGCTTGGGTGCGGAAACCGCTTTCGATCGGCGGACGTC 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 SMSAHRCSKCTACVTSBTSBGSHKVRGMVSVVVBHDHATSKYGGHTGYGCSGSGTG 180

Qy 553 TTCAATCACAAGATTCGATCTGCTGTTGAAGGGCAACGATATCATCTTCGCGCTTGGCTT 512
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 HTDCCBACSAVBDSSCBWAYCTGCTBRARKVNSTVTYRSSBMYNSBKRBVSRYVM 240

Qy 613 GCGCTGCGGATCCGACACAGCTTAGCGTTCGAGGTTGCTCGGCGCAAGCCCAAGGT 672
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 MYRCCSVSVATCCVMSRINTGGARHTKWCNCTWCRFCRMCRCRHSRSSAMYKGG 300

Qy 673 GTTACGATCTGCTGCGAGCAATCTGGTGC--CGGTTTCGAAGGATGATCGAAGGCA 730
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 CSRTGMVKWRTYTSKSSDVNRBCSTSVARWSMGSDSVSVSSGNGCVKRBVNTSR 360

Qy 731 AACCTGCAATCCATGGAACATGGCAAGCGCGCTGCGCATCAGCATATCGATCTGGC 790
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 TSGSVRSYSHCRBBSRSTAGWYCGSYKWTWCVDNMKNSANCGBRTVMSBHSANC 420

Qy 791 TGGCGGCGTACTCAAGTCAAGCTGCAACCGTGGCGGATCCCGAGCTCAGCGATTCGAATGT 850
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 GVSYVGBYCGSYBMSVWVHVRBRBVRDMHRTVSTSGCSTSYVGGAYCRVBADCRMD 480

Qy 851 ACGTGGCGGTCTGATGACACGACCCACC 877
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 481 SCVYRNSVHBYKYWSCKYRSMHYC 507

RESULT 5
US-10-259-194A-593
; Sequence 593, Application US/10259194A
; Publication No. US2004010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghassemian, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP

; CURRENT APPLICATION NUMBER: US/10/259,194A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 593
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Zea mays
US-10-259-194A-593

Query Match      3.8%; Score 43.2; DB 16; Length 498;
Best Local Similarity 46.6%; Pred. No. 0.0013;
Matches 138; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy 698 CGGTGCGGGTTTTCGAAGGATGATCGAAGCAACCTGCAATCCATGCGCAACATTGGCA 757
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 200 CGGAGGCGGGTCCACCGCCAGCAGCGCGGAGCTGGCCCGCTAGCGGCGGACG 259

Qy 758 GCAAGCGCGTCCGCAATCAGCATATCGATCTGGTTCGCGGCGTACTCAAGTCAACCGT 817
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 GCGACGCGGGGCTCAGCGTGGACTACGGCTCGCTCGCCGACATCGACGCTTCTTCCAGT 319

Qy 818 GSCCGGATCCCGACTCAGCGAGTTCGATGTTGTTGTTGCGGTGATGAGACCGACCC 877
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 GCGCCAAAGTGCATGAGCAGCAGCCACTACTACTAGCAGCGCTCATGACCCCTCGCGG 379

Qy 878 TCTACTTCCAGACGCTGGGCAAGTCGTGAGCTCAAGAGGAGCGGCGGACTCTTTCGAGC 937
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 CTTTCTTCGCGCGCGCCATCGGTTCGCGGCGGACGACGCGCGCGCGGAT 439

Qy 938 GAGATTCACGAAAATGGTAGGCTTCGGCTTACGGCTTACGGCTTCAATGATGACAC 993
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 440 GCTGGGAGGACGCGCAGCGCATTCGGCTCTGGAGCTTCAGCTTCCCGCTCTGGAC 495

RESULT 6
US-10-282-122A-26243
; Sequence 26243, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
```

PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: 60/269,308
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 26243
LENGTH: 2823
TYPE: DNA
ORGANISM: Mycobacterium bovis
US-10-282-122A-26243

Query Match 3.6%; Score 40.6; DB 13; Length 2823;
Best Local Similarity 49.8%; Pred. No. 0.019;
Matches 103; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 346 ATCCGCGCGCGCTTTGAAGACGTTCCCGGTTGAAGAGGCCAAAGGTCTTATCTTCGTT 405
Db |||||
QY 406 TACGTAGCGGAGCGGCAACCAACGCGCTTATCGAAGATGTCGCGCGCTTCCTTGAT 465
Db |||||
QY 727 GTGCGCGTCCGCGCGGCTTTGTCGCGCTGACGCTGATCGCGCGCGGAGATCGGC 786
Db |||||
QY 466 GAAACCGCGCCATTCACGGCCAAACATCGGCTCGTGGCTCGAACTGCGGCTTGGGTGCG 525
Db |||||
QY 787 GCTGACGTGCGCTTTGGCACCACAAACGGTTCCGATGCGGCTTGGCGGCCG 846
Db |||||

QY 526 GAAACCGCGCTTTGATCGGGGCGACGTC 552
Db |||||

QY 847 CATGCGGGTACCTTGGCGTGCACGCC 873
Db |||||

RESULT 7
US-10-282-122A-28428
Sequence 28428, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELTRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: 60/269,308
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 28428
LENGTH: 2826
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28428

Query Match 3.6%; Score 40.6; DB 13; Length 2826;
Best Local Similarity 49.8%; Pred. No. 0.019;
Matches 103; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 346 ATCCGCGCGCGCTTTGAAGACGTTCCCGGTTGAAGAGGCCAAAGGTCTTATCTTCGTT 405
Db |||||
QY 406 TACGTAGCGGAGCGGCAACCAACGCGCTTATCGAAGATGTCGCGCGCTTCCTTGAT 465
Db |||||
QY 727 GTGCGCGTCCGCGCGGCTTTGTCGCGCTGACGCTGATCGCGCGCGGAGATCGGC 786
Db |||||
QY 466 GAAACCGCGCCATTCACGGCCAAACATCGGCTCGTGGCTCGAACTGCGGCTTGGGTGCG 525
Db |||||
QY 787 GCTGACGTGCGCTTTGGCACCACAAACGGTTCCGATGCGGCTTGGCGGCCG 846
Db |||||

QY 526 GAAACCGCGCTTTGATCGGGGCGACGTC 552
Db |||||

QY 847 CATGCGGGTACCTTGGCGTGCACGCC 873
Db |||||

RESULT 8
US-10-425-114-20746/c
Sequence 20746, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(5313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 20746
LENGTH: 1801
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3245-118-C5_FLI
US-10-425-114-20746

Query Match 3.6%; Score 40.4; DB 13; Length 1801;
Best Local Similarity 51.1%; Pred. No. 0.019;
Matches 95; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 372 CCGGTTGAAGAGGCCAAAGGTCTTATCTTCGTTTACGTAGCGGCGGGAACCAACGCC 431
Db |||||
QY 668 CCGGATGTCGGGACCAACGCGCTCGCGCTTCGCGCGCGGACCAACGCGGCGCC 609
QY 432 GCTTATCGAAGATGTCGCGCGCGGCTTCCTTGATGAAACCGCGGCTTATCGGCAACA 491
Db |||||

QY 492 TCGGCTCGTGGCTCGAACTGGCTTGGTGGGAAACGGCTTTGATCGGGGACGT 551
Db 548 TCTCGTGGTAGCGCGCGACAGCACGGTGGTGGCGAACTCCGGCTTGATCTCGAACTCG 489
QY 552 CTTTCAT 557
Db 488 CGGCGT 483

RESULT 9
US-10-282-122A-25435
; Sequence 25435, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELIWA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25435
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Mycobacterium avium
US-10-282-122A-25435

Query Match 3.5%; Score 40.2; DB 13; Length 1830;
Best Local Similarity 47.4%; Pred. No. 0.022;
Matches 120; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 688 GCGAGCAATTCGGTCCGGTTTTCGAAGGCATGATCGAAGCAAACTGCAATCCATGGC 747
Db 1306 GACGAGTGGCGTGGCGGTTCATCGACCGCGGTGTGGTGGTCATCCGAAACCAACGTC 1365
QY 748 AACTTGGCAGCAAGCGGTGGCATCGATCATCGATCTGGTGGCGGCGTACTCAAG 807
Db 1366 AACCTGGCTGGCGTGGCGGATCATCGGTTCACGCTTCAACGTGGCGCGGAGGCAAG 1425
QY 808 GTCGAACCGTGGCGGATCCGAGCTCACGAGTTCAATGTTACGTACGTGCGGTGATGAG 867

Db 1426 GCCACCGAGTGGCCCAACCGGAGGGGTGGAGATCCGTACTACTCGGTGATCTACCAG 1485
QY 868 ACAGGACCACTCTACTTCCAGAGCGTGGGCAAAAGTGTGACGTCAAAAGGAGCGGACAG 927
Db 1486 GCGATCGACGAGATCGAGAAGGCCCTCGCGCGCATGCTCAAGCCGATCTTACGAGGAGAC 1545
QY 928 TCCTTCGAGCGAG 940
Db 1546 CAGCTGGGCGCGG 1558

RESULT 10
US-10-369-493-31790
; Sequence 31790, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31790
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Thermobifida fusca
US-10-369-493-31790

Query Match 3.5%; Score 40; DB 16; Length 1185;
Best Local Similarity 54.9%; Pred. No. 0.022;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 754 GCGAGCAAGCGGTGGCGCATCGATATCGATCTGGTGGCGGCTACTCAAGTGGAA 813
Db 988 GCCAACAAAGCACGTCAACCGGCGATCCGGATGGTGTGTCACGCGCGCGCGGTGTC 1047
QY 814 CCGTGGCGGATCCCGAGCTCACGAGTTCGAAATGGTACGTGCGGTGATGAGACAGC 873
Db 1048 ACCGTGGCGGAGGAGGCTCCCTCTACGACTACCACTGGCCACCTAGCAGCGGC 1107
QY 874 CACTCTACTTCCAGAGCGTGGC 897
Db 1108 GACACCTACGACGAGCGCTGGCC 1131

RESULT 11
US-09-758-759-130
; Sequence 130, Application US/09758759
; Publication No. US20040101832A1
; GENERAL INFORMATION:
; APPLICANT: Hosted, Thomas J.
; APPLICANT: Wang, Tim X.
; APPLICANT: Hoxan, Ann C.
; TITLE OF INVENTION: Evernimicin Biosynthetic Genes
; FILE REFERENCE: ID0983K US
; CURRENT APPLICATION NUMBER: US/09/758,759
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/175,751
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 130
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Micromonospora carbonacea

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(606)
; OTHER INFORMATION: evbu
US-09-758-759-130

Query Match      3.5%; Score 39.8; DB 12; Length 606;
Best Local Similarity 45.4%; Pred. No. 0.019;
Matches 143; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

Qy 772 ATCAGCATATCGATCTGGCTGCGGGCGTACTCAAGGTCGAACCGTGGCCGGATCCCGAG 831
Db 10 ATCAGCTCGACCTCGGATGGCCGCCGCTCTGTATCGAGCTGGCCCTCGCCGGCGG 69
Qy 832 CTCAGCGAGTTGATGTTAGTGTGCGGCTGCGATGAGACACCGACCTTATTTCCAGACG 891
Db 70 GTGCGGTACGTGATGGAACCTGTGTGTCAGGACCCCGGCCACCGGGGAACCGATC 129
Qy 892 CTGGGCAAGTCTGACGTCAAGGAAGCGGAGACTCTTCGAGCGAGATTCCACGAA 951
Db 130 CTGGACGAGTCTCTGGCGGGATGCGCGCGGACACCGCGACCTCTCTCTGGGTG 189
Qy 952 AAATGGGTAGGCTTCGCGCTTAACCGCTTCAATGATGACGACATCATGCGACGTGAATCG 1011
Db 190 CAGCGGCTGCGCACGGGCTGCGGACCGGATCTCTCGGCGACCTGGTGGCCGGGGGTG 249
Qy 1012 ATGGAGCGCTTCTACGCTGATGATCGCGTGTGCTCGGAGAAATCTGTCGACCGGAC 1071
Db 250 GTGCGGAGCTGAGGAGACCGAGTGGCTTCATCTCAGCTGACCGCTATCCGACGGCG 309
Qy 1072 CGCGCAATCATCGAG 1086
Db 310 GACGCCACCGTGGAG 324

RESULT 12
US-09-758-759-1/c
; Sequence 1, Application US/09758759
; Publication No. US20040101832A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tim X.
; APPLICANT: Horan, Ann C.
; TITLE OF INVENTION: Everninomicin Biosynthetic Genes
; FILE REFERENCE: ID0983K US
; CURRENT APPLICATION NUMBER: US/09/758,759
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/175,751
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 109519
; TYPE: DNA
; ORGANISM: Micromonospora carbonacea
US-09-758-759-1

Query Match      3.5%; Score 39.8; DB 12; Length 109519;
Best Local Similarity 45.4%; Pred. No. 0.16;
Matches 143; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

Qy 772 ATCAGCATATCGATCTGGCTGCGGGCGTACTCAAGGTCGAACCGTGGCCGGATCCCGAG 831
Db 83879 ATCAGCTCGACCTCGGATGGCCGCCGCTCTGTATCGAGCTGGCCCTCGCCGGCGG 83820
Qy 832 CTCAGCGAGTTGATGTTAGTGTGCGGCTGCGATGAGACACCGACCTTATTTCCAGACG 891
Db 83819 GTGCGGTACGTGATGGAACCTGTGTGTCAGGACCCCGGCCACCGGGGAACCGATC 83760
Qy 892 CTGGGCAAGTCTGACGTCAAGGAAGCGGAGACTCTTCGAGCGAGATTTCACGAA 951
Db 83759 CTGGACGAGTCTGGCGGGATGCGCGCGACACCGCGCACACCCGCTCTCTCTGGGTG 83700

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(606)
; OTHER INFORMATION: evbu
US-10-282-122A-26553

Query Match      3.4%; Score 39.2; DB 13; Length 1308;
Best Local Similarity 45.7%; Pred. No. 0.042;
Matches 137; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Qy 641 GTTCCCGAGGTGTGTCGGGGCAAGCCCAAGAGTGTTCACGATCTGCTTGGCGACATTCCG 700
Db 737 GTACCGTCGAGGCGCTGGAAGAGCCCTGATGGGTATCCAGGTGGACGAGGTGGTGC 796
Qy 701 TGGCGGTTTCGAGGACATGATCGAAGGCAACCTGCAATCTTCATCGCAATTCGACGCA 760
```

Db 797 TCGGGGTGATCGACCGGGTGTCCGGCGGATACCGAARACCAACGTCACACCTGGCGTCGG 856
Qy 761 AGCGCTCGCCATCAGCATATCGATCTGGTCCGGCGGTACTCAAGTTCGAACCGTGGC 820
Db 857 CTTCCGATGGGTGATCATCGGTTTCAATGTGCGCGCCGAAAGCAAGCGGACCGAGCTGG 916
Qy 821 CGGATCCGAGCTCAGCAGTTCGAATGGTACGTCCGGTTCGATGAGACAGCACCTCT 880
Db 917 CGAGCCGGAAGGGTGGAGATCGCTACTACTCTCGGTCACTACCAAGGGATCGACGAGA 976
Qy 881 ACTTCCAGACGCTGGGCAAAAGTCTGTGACGTCAAGGAAGCGGAGACTCTCTTCGAGCGAG 940
Db 977 TCGAGCAGCGCTCGTGGTGGCTGTCTCAAGCGCATCTACGAGGAAACCCAGCTGGTGGTGG 1036

RESULT 14

US-10-282-122A-28597
; Sequence 28597, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28597
; LENGTH: 2703
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28597

Query Match 3.4%; Score 39.2; DB 13; Length 2703;
Best Local Similarity 45.7%; Pred No. 0, 056;
Matches 137; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Qy 641 GTTCCGAGTTGCTCGCGGCAAGCCAAAGTGTTTACATCTGCTGGCGAGATTCGG 700
Db 2129 GTACGCTCAGGCGCTGGAAGAGGCCCTGATGGGTATCCAGGTGACACGAGGTGTGTC 2188

Qy 701 TCGCGGTTTTTGAAGGCATGATCGAAGGCAAACTGCAATCCATGGCAACATTGGCAGCA 760
Db 2189 TCGCGGTGATCGACCGGGTGTTCGGCGGCATTACCGAACAACACTCAACTGGCGTCGG 2248
Qy 761 AGCGCTCGCCATCAGCATATCGATCTGGTCCGGCGGTACTCAAGTTCGAACCGTGGC 820
Db 2249 CTTCCGATGGGTGATCATCGGTTTCAATGTGCGCGCCGAAAGCAAGCGGACCGAGCTGG 2308
Qy 821 CGGATCCGAGCTCAGCAGTTCGAATGGTACGTCCGGTTCGATGAGACAGCACCTCT 880
Db 2309 CGAGCCGGAAGGGTGGAGATCGCTACTACTCTCGGTCACTACCAAGGGATCGACGAGA 2368
Qy 881 ACTTCCAGACGCTGGGCAAAAGTCTGTGACGTCAAGGAAGCGGAGACTCTCTTCGAGCGAG 940
Db 2369 TCGAGCAGCGCTCGTGGTGGCTGTCTCAAGCGATCTACGAGGAAACCCAGCTGGTGGTGG 2428

RESULT 15

US-10-338-110-119/c
; Sequence 119, Application US/10338110
; Publication No. US20040023254A1
; GENERAL INFORMATION:
; APPLICANT: Fuhrmann, Jeffrey J.
; APPLICANT: Ronesser, James A.
; TITLE OF INVENTION: A Method To Assess Quorum Sensing Potential Of Microbial
; TITLE OF INVENTION: Communities
; FILE REFERENCE: HER-0056
; CURRENT APPLICATION NUMBER: US/10/338,110
; CURRENT FILING DATE: 2003-01-07
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 119
; LENGTH: 536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus Sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7)..(7)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (21)..(21)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (33)..(33)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (69)..(69)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (87)..(87)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (213)..(213)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (226)..(226)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (261)..(261)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (273)..(273)
; OTHER INFORMATION: n is a, c, g, or t

```

FEATURE:
; NAME/KEY: misc feature
; LOCATION: (347)..(347)
; OTHER INFORMATION: n is a, c, g, or t
FEATURE:
; NAME/KEY: misc feature
; LOCATION: (357)..(357)
; OTHER INFORMATION: n is a, c, g, or t
FEATURE:
; NAME/KEY: misc feature
; LOCATION: (398)..(398)
; OTHER INFORMATION: n is a, c, g, or t
FEATURE:
; NAME/KEY: misc feature
; LOCATION: (419)..(419)
; OTHER INFORMATION: n is a, c, g, or t
FEATURE:
; NAME/KEY: misc feature
; LOCATION: (509)..(509)
; OTHER INFORMATION: n is a, c, g, or t
FEATURE:
; NAME/KEY: misc feature
; LOCATION: (517)..(517)
; OTHER INFORMATION: n is a, c, g, or t
US-10-338-110-119

```

```

Query Match      3.4%; Score 39; DB 17; Length 536;
Best Local Similarity 15.2%; Pred. NO. 0.034;
Matches 36; Conservative 111; Mismatches 90; Indels 0; Gaps 0;

QY 568 TCGATCCTCGTGAAGGCAACGATCATCTGCGCGTTGGCTTGGCGCTGGCGATCCC 627
Db 422 BCGNTSSDVSHKBAVYCCGWTSMKHGBGWMRSCBNCATSSVYVDWRSRYRYSBC 363
QY 628 GACCACTTACCGTTCCGAGTTGCTGCGGCAAGCCCAAGGTTTACGATCTGCTT 687
Db 362 SAYSANSBVMBGNCSSVBSRSSHGCKSWYTBASGVYKBHSSMSARAYWMBKAY 303
QY 688 GCGGACATCGGTGCGCGTTTTCGAGGATGATCGAGGCAACCTCCATCCATGCC 747
Db 302 SGCCMRKTSGGSDYGYGKGYGWAACGNGWADYTCCANAYSKMBGGATBSRSGGY 243
QY 748 AACATTGGCAGCAAGCGCTGCGCATCAGCATATCGATCTGGCTGCGGGCGTACTC 804
Db 242 RKKBYSBYGVMBVSNRKVSSYRABASBNMYTVAGCAGRTWVGSSHVBGTSGTV 186

```

Search completed: June 19, 2004, 09:16:40
Job time : 546.885 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2004, 04:04:20 ; Search time 1331.47 Seconds
(without alignments)
10742.394 Million cell updates/sec

Title: US-10-658-691-2
Perfect score: 330
Sequence: 1 ataccgcaaggtccgcgt.....tgcggaagcacaagtcataga 330

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 5940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_ov.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_btgo_hum.*
40: em_btgo_mus.*
41: em_btgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	330	100.0	4154	1	AF442494	AF442494 Sphingomo
2	330	100.0	13651	1	AB095953	AB095953 Sphingomo
3	115.2	34.9	12376	1	AE014404	AE014404 Brucella
C 4	115.2	34.9	12429	1	AE009536	AE009536 Brucella
C 5	114.2	34.6	345897	1	AP002395	AP002395 Mesorhizo
C 6	114	34.5	300140	1	AP005956	AP005956 Bradyrhiz
C 7	107	32.4	1351	1	CFRDXB	XS1607 C.crescentu
C 8	107	32.4	10158	1	AE006011	AE006011 Caulobact
C 9	106.4	32.2	4151	1	RCFDXE	Y11304 R.capsulatu
C 10	105.6	32.0	6407	1	AB121977	AB121977 Xanthobac
C 11	100.6	30.5	349746	1	BX572605	BX572605 Rhodosphe
C 12	89	27.0	309350	1	AP005215	AP005215 Corynebac
C 13	86.2	26.1	10495	1	AE008062	AE008062 Agrobacte
C 14	86.2	26.1	10883	1	AE009097	AE009097 Agrobacte
C 15	86.2	26.1	306250	1	SME591788	ALS91788 Sinorhizo
C 16	85.8	26.0	299450	1	AP005938	AP005938 Bradyrhiz
C 17	85.2	25.8	6040	1	AB062679	AB062679 Bradyrhiz
C 18	84.4	25.6	4379	1	ASP311718	AJ311718 Acinetoba
C 19	81.2	24.6	347894	1	BX640431	BX640431 Bordetell
C 20	81.2	24.6	348542	1	BX640446	BX640446 Bordetell
C 21	79.6	24.1	348934	1	BX640417	BX640417 Bordetell
C 22	79.2	24.0	349497	1	BX640440	BX640440 Bordetell
C 23	74	22.4	4398	6	AX645690	AX645690 Sequence
C 24	74	22.4	4398	6	AX661050	AX661050 Sequence
C 25	74	22.4	8766	6	AX645680	AX645680 Sequence
C 26	74	22.4	8766	6	AX661040	AX661040 Sequence
C 27	74	22.4	12342	6	AX645679	AX645679 Sequence
C 28	74	22.4	12342	6	AX661039	AX661039 Sequence
C 29	74	22.4	23656	1	AF333761	AF333761 Rhodococc
C 30	74	22.4	23656	6	AX645678	AX645678 Sequence
C 31	74	22.4	23656	6	AX661038	AX661038 Sequence
C 32	73.6	22.3	321	6	AX661046	AX661046 Sequence
C 33	73.6	22.3	321	6	AX661130	U17130 Rhodococcus
C 34	72.8	22.1	6458	2	AC079815	AC079815 Trypanoso
C 35	68	20.6	146584	2	AC079815	AC079815 Trypanoso
C 36	67.2	20.4	4416	1	AF096864	AF096864 Pseudomon
C 37	65.6	19.9	306	6	AX120694	AX120694 Sequence
C 38	65.6	19.9	306	6	BD162811	BD162811 Novel pol
C 39	65.6	19.9	318	6	AX120696	AX120696 Sequence
C 40	65.6	19.9	318	6	BD162813	BD162813 Novel pol
C 41	65.6	19.9	441	6	AX065549	AX065549 Sequence
C 42	65.6	19.9	328050	1	AP005275	AP005275 Corynebac
C 43	65.6	19.9	349880	6	AX127144	AX127144 Sequence
C 44	64.8	19.6	2497	3	DMHSPG3	X06542 Drosophila
C 45	64.4	19.5	322774	3	CEY73F8A	AL132862 Caenorhab

ALIGNMENTS

RESULT 1
AF442494
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AF442494
Sphingomonas sp. GTIN11
AF442494
AF442494.1
GI:17227030
Sphingomonas sp. GTIN11
Sphingomonas sp. GTIN11
Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
Sphingomonadaceae; Sphingomonas.
1 (bases 1 to 4154)
Kilbane,J.J., II, Daram,A., Abbasian,J. and Kayser,K.J.
Isolation and characterization of Sphingomonas sp. GTIN11 capable
of carbazole metabolism in petroleum

4154 bp DNA linear BCT 22-OCT-2002
Sphingomonas sp. GTIN11 carbazole operon, partial sequence.

Pred. No. is the number of results predicted by chance to have a

JOURNAL Biochem. Biophys. Res. Commun. 297 (2), 242-248 (2002)
MEDLINE 2222130
PUBMED 12237109
REFERENCE 2 (bases 1 to 4154)
AUTHORS Kayser K.J., Daram A., Abassian, J. and Kilbane, J.J. III.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2001) Biotechnology, Gas Technology Institute,
1700 South Mount Prospect Road, Des Plaines, IL 60018, USA
FEATURES
source
1. .4154
/organism="Shingomonas sp. GTIN11"
/mol_type="genomic DNA"
/db_xref="taxon:174635"
-1. .4154
/gene="carbazole operon"
650. .1786
/gene="carAa"
650. .1786
/gene="carBa"
/codon_start=1
/transl_table=11
/product="carbazole dioxygenase"
/protein_id="AAL37978.1"
/db_xref="GI:17227031"
/translation="WAPSPSTAEETKWEPIYRAKLGFRNHWYFVRLASBIAGTPV
PVKLQCEKILLNRVGGKVYAIQORCLHRGVTLSDRVCYSKNTISCMVHGWTYRWDG
RVLDILTPSGVOIGRRALKTPVEAEAGLIPVYVDGEPTLIEDVPPGLDENRAI
HGQRLVANSNLTGAENGDFDAGHFIHKNSILVKGNDIILPLGAPGDDOLTRSEVA
ACKPGVYDLLSEHSPVPEGMIEGKPAHNGISKRVASISLWLPVLKVEPWPDP
EUTQFVYFVDETSHLIFQTLGKVVTSKEADSFERSFHEKVGWGLAUNGFNDDIIMA
RSMGPFYADDDSEELFEPDRAIIEMRGUASQHNRIQEAR"
1735. .2067
/gene="carBa"
1735. .2067
/gene="carBa"
/note="meta cleavage enzyme subunit"
/codon_start=1
/transl_table=11
/product="2-aminobiphenyl-2,3-diol 1,2-dioxygenase"
/protein_id="AAL37977.1"
/db_xref="GI:17227032"
/translation="WAGCQSAQSRHSGLTGGTTASYPVNRMLMOELFTNPQNVLEFR
ADREALYERYGLSSAQRALDEGGFALTAVLGHFVQLMHFMLNPNAPDFVSVKAY
RKMVDNRG"
2060. .2863
/gene="carBb"
2060. .2863
/gene="carBb"
/note="meta cleavage enzyme catalytic subunit"
/codon_start=1
/transl_table=11
/product="2-aminobiphenyl-2,3-diol 1,2-dioxygenase"
/protein_id="AAL37978.1"
/db_xref="GI:17227033"
/translation="WAEIVGVFATSHLVFGSPDGAQALRVVVEGNEIGRRVRLRPD
LTIIVGSDHFNITRLQPFVGVSDTFTPLGMDIQRFPAGNRAFAESLCRAAD
RFDLAGEBLRDPHGVNMTDPDGRIPVPIVYNNMTPTPTAARAOLGDIIVAE
AVGLDLPRLHVVATGSLWINIPGHGEVNAFEDSRVIAAFQSGDMRWLRALDTE
TLKXNAGSLRIVNWMAAATLPGRRAEKIYEPNPQMGMGGIAIV"
2906. .3730
/gene="carC"
2906. .3730
/gene="carC"
/codon_start=1
/transl_table=11
/product="meta cleavage compound hydrolase"
/protein_id="AAL37979.1"
/db_xref="GI:17227034"
/translation="WVLAATAGRSVTVRGMEYTHYEQSGDVVVLVHGGAGADSMGN
WRGMPVLAADRVYVADMLGFTAKPADPFPVQAPARTDHLAFLDALGLSNVAV
GNSMGFALGVAVERPGLVRLNMGSLVSKIDPALEPVLGVDFTRGMRILVRA
LTTNFGQIDAMDIVYALVDPETERAYSATQWIRDOGGLYVEDDVIIRITAPTLI
VNGKLDKVLNAYKFLLEIGPSMGYNPDGCHWAMIEHPVDFPARTTAATFAAAQ"

gene 3771. .4100
/gene="carAc"
3771. .4100
/gene="carAc"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="AAL37980.1"
/db_xref="GI:17227035"
/translation="MTAKRVVIFRAAGGFHELVETEAGVSLMEAAVINGVDGIEAVCG
GACACATCHVYVGPWMLDKPSETDEMDLCAERAPHSRLSCQIRLTLDLDGLTL
ELPKAQS"
ORIGIN
Query Match 100.0%; Score 330; DB 1; Length 4154;
Best Local Similarity 100.0%; Pred. No. 9.3e-56;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGACCGCAAAGGTCCGCGTGTATCTTCGCGCAGCGCGGGCTTCGAGCATCTGTCGAA 60
Db 3771 ATGACCGCAAAGGTCCGCGTGTATCTTCGCGCAGCGCGGGCTTCGAGCATCTGTCGAA 3830
Qy 61 ACCGAAGCGGAGTATCGCTCATGGAAGCGCGCTTCTGAACGGCGTGGACGGTATCGAA 120
Db 3831 ACCGAAGCGGAGTATCGCTCATGGAAGCGCGCTTCTGAACGGCGTGGACGGTATCGAA 3890
Qy 121 GCCTTTTCGCGGGCGCTGTGCTGTGCGCACGTGCACGTTTACGTTGCGCCCGAGTGG 180
Db 3891 GCCTTTTCGCGGGCGCTGTGCTGTGCGCACGTGCACGTTTACGTTGCGCCCGAGTGG 3950
Qy 181 CTAGATCGCTGAAAACCGCGAGTAGACCGAAGAGAAATGTCATGCTGCGTAGCGGAA 240
Db 3951 CTAGATCGCTGAAAACCGCGAGTAGACCGAAGAGAAATGTCATGCTGCGTAGCGGAA 4010
Qy 241 CGTGGCGCATTCGCGGCTGTCTGCGAGATCCGCTTACCGACCTGCTCGACGCGCTG 300
Db 4011 CGTGGCGCATTCGCGGCTGTCTGCGAGATCCGCTTACCGACCTGCTCGACGCGCTG 4070
Qy 301 ACCCTGGAATGCGGAAGGCACAGTCATGA 330
Db 4071 ACCCTGGAATGCGGAAGGCACAGTCATGA 4100
RESULT 2
AB095953
LOCUS AB095953 13651 bp DNA linear BCT 04-FEB-2003
DEFINITION Shingomonas sp. KAI car operons, complete cds.
ACCESSION AB095953
VERSION AB095953.1 GI:28201219
KEYWORDS
SOURCE Shingomonas sp. KAI
ORGANISM Shingomonas sp. KAI
Bacteria; Proteobacteria; Alphaproteobacteria; Shingomonadales;
Shingomonadaceae; Shingomonas.
REFERENCE 1
AUTHORS Inoue, K., Nojiri, H., Nakai, S., Endoh, T., Urata, M., Ashikawa, Y.,
Saiki, Y., Yoshida, T., Habe, H. and Omori, T.
TITLE Divergent structures of carbazole-degrading car operons isolated
from several gram-negative bacteria
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 13651)
AUTHORS Nojiri, H. and Inoue, K.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-2002) Hideaki Nojiri, Biotechnology Research
Center, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo
113-8657, Japan (E-mail:anojiri@mail.ecc.u-tokyo.ac.jp,
Tel:81-3-5841-3067, Fax:81-3-5841-8030)
FEATURES
source
1. .13651
/organism="Shingomonas sp. KAI"
/mol_type="genomic DNA"
/strain="KAI"
/db_xref="taxon:164608"

```
gene
complement(127. .2199)
/gene="ORF10"
CDS
complement(127. .2199)
/gene="ORF10"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56754.1"
/db_xref="GI:28201220"
/translation="MISEKIGPOHLERKAILYVQSSVHQVHLNHRSSITLOVAMRGL
TAMGWSHETIDDLGRSAAGVTRAGFDKRWAEVCLGKGVAAAEVSRFARNSRDW
QQIEMCRVDTVIDOETVYAPQGNDRLLGLKGLNEYEDLLRQSLRAREKA
RREGELVVAAPGVGADRIEKDPDRVQBAIALVDFKVAELGSAQCALLWFTEHGLD
LPKRNNGDVVWRPSTVATHTMIENPIYGGAYAGKSAATGYDGTAMRSRKRAR
KLWYALTPGSHGVSWERSEDIKMVS DNVPSTSHHGAPKGDLAGLIRCRGCR
DLVYATGAKHDIPRYCSGRLDNGSPRCIAFGGLRVDDAIERALLQVLEPGAIAS
VEAQAADRDVDVDMEDLEAAYADRAEROYDAADPNELVAELEFWFWEAL
TRAGEVAVIADASTALHPALFSLKIDGLADSLAEQVWNAPOS DAKLRIVRTLIQ
EVVADHDASEIVLLHWGVLHTDLRPRRKGQNSTADIIAAVRLVLIANDD
LIAGILNRNLGVTGHGNNRTRERTALRSRHRIPVFTVADGPAPWLNLQSAARHTGV
ASKTLRIAAAGBIKGIHPLPEGFWIFCRTELDGSAAHHLAKARQNPKYPTGSHPDQ
QTLFSSTT"
2502. .2852
/gene="ORF9"
2502. .2852
/gene="ORF9"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56755.1"
/db_xref="GI:28201221"
/translation="MGRTHPCQCARSRVHEPRNLEPPVRSRSLATAQINRTGFD
LMDVDMPASGPMFAIENFNGFVGVPSCCTTARTAVLAIGHRHQRRHRIRP
FVPLRSTSVIRMR"
3162. .3785
/gene="ORF8"
3162. .3785
/gene="ORF8"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56756.1"
/db_xref="GI:28201222"
/translation="MKYNAAERFHREYGTIKSIKTPNFIVHTLOEVNDGLAHGRY
FMDVADGIALQSDSDVHTPKPTPKDALKMAREYFDEWPPSAKRNIAKFDVG
QGYLRDAADFQHATERLYHCVLVCTFYFPHVNTLGFURSQANLIDRLMYVWPDN
RKQRMPEKLKQAYKARYSKYRISEELTWLGEQVEELGRVYHAV"
3862. .4188
/gene="ORF7"
3862. .4188
/gene="ORF7"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56757.1"
/db_xref="GI:28201223"
/translation="MQRDSGDSRQNGRSATKNRDILLAPVSCANAGFRKREAGEGS
NFGFGQREIIDIYAEVSDCILDVGNAEQDLYGAEIACGFVNQRRLRSRVRATFP
TVETNR"
complement(4409. .5089)
/gene="carR"
complement(4409. .5089)
/gene="carR"
/codon_start=1
/transl_table=11
/product="transcriptional regulator of car operon"
/protein_id="BAC56758.1"
/db_xref="GI:28201224"
/translation="MISTTWTSLDKESPAHRAAAFTIRGAILEARFAPQRLVESELS

SLTGLGRGPVREARLLAAEGLVAIERNKGAIVARASKAMMVEFVEVRELLGLEAARR
AALRAATSPHLESIERALDEESLENVDDSGIILMEANEHLHLQVLVDAAESVIERVLV
OMQLPELSSIFFRMLSPTIWRSHREDHIGILNTILKGDGEAAEQMRHARVHTAALFA
DLPDELLR"
5192. .6328
/gene="carAa"
5192. .6328
/gene="carAa"
/codon_start=1
/transl_table=11
/product="terminal oxygenase component of CAR
1,9a-dioxygenase"
/protein_id="BAC56759.1"
/db_xref="GI:28201225"
/translation="MANQPSIAERTKWEPIYIRAKLGRFNHWYVPIRLASIAEGTPV
PVKILGKILINRVGGKYVIAQDRCLARGVTLSRVCEYSKNTISCWYHWTYRWDDG
RLVDILTVPQSVQIRALKTFPVVEAKGLIFVVVGCEPTPLIEDVPPGLDNRRI
HGOHLVASNWLGAENGFDAGHVIHKNSILVAKNDIILPLGAPGDDOLTRSEVA
AKQKGVYDILLGEHSVPVPEGNIEGKPAHNGSKRVAISISINLPGVLKVEWPPDP
ELTQFEWYVVDVETSHLYFQTLGKVTSKEAASFEFFHEKWKVGLALNGFNDDIIMA
RESMEPFYADDRGWSEELLFEPDRAIIEMRWGLASQHNRIQEAR"
6328. .6609
/gene="carBa"
6328. .6609
/gene="carBa"
/codon_start=1
/transl_table=11
/product="subunit of meta cleavage enzyme"
/protein_id="BAC56760.1"
/db_xref="GI:28201226"
/translation="WGTATSYVNRLMQELFTNPQNVLPFRADREALYERYGLSSAQR
AALDEGGGALTAVGLHPLVLMHFMLTNPAPDFSVKAYRKWDRNG"
6602. .7405
/gene="carBb"
6602. .7405
/gene="carBb"
/codon_start=1
/transl_table=11
/product="subunit of meta cleavage enzyme"
/protein_id="BAC56761.1"
/db_xref="GI:28201227"
/translation="MAEIVGVFATSHVLFSGPDGDAQALRVVEGNEIGRRVRLLRPD
LLVIGSDHFNITTRLOPFTVGSVSTFTFGMDIEQRFAGNRAFAESLCARAAD
RFDLAQGSELPRDGVVPLMIDPDGRIIPVPIYVNNMTPTTAAARAAGLQIVAE
AVGLDPSHLRVVVVATGTLGSHWINIPGHGEVNAEFORRVIAAFQSGDMRWLRADTE
TLLKNAGNGGLEIVNWMAATLFGRRAEKIYBEPQWMTGMGIAIV"
7448. .8272
/gene="carC"
7448. .8272
/gene="carC"
/codon_start=1
/transl_table=11
/product="meta cleavage compound hydrolase"
/protein_id="BAC56762.1"
/db_xref="GI:28201228"
/translation="WVLAATAGRSVTVRGMEITYHEQSGDVVVLVHGGAGAGSMGN
WRGVPLADRYVIAVDMLGTAKPADFPVFSQAARTDHLGFLDALGSLNVALV
GNSMGASALGVAVRPGVRLKVLMSGAGLVSKIDPALPEVLGYDFTREGMIRLVA
LTTDNFQIDDAVIDYRVALSVDPTRRAYSATMOWIQOGLYVEDDYIRITAPTLI
VNGKLDKVPVPIANAYKFLLELIGPSWGIYMPDCGHWAMIEHPVDVFARTTAAFEAAQ"
8313. .8642
/gene="carAc"
8313. .8642
/gene="carAc"
/codon_start=1
/transl_table=11
/product="ferredoxin component of CAR 1,9a-dioxygenase"
/protein_id="BAC56763.1"
/db_xref="GI:28201229"
/translation="MTAKRVIFRAAGFEHLVTEAGVSLMEAAVNGVDGIRAVCG
GACACATCVTVGPEWLDALKPPSETEDEMLDCVAERAPHSRSLSCQIRLTLDDGLTL
ELPQAQS"
8684. .10999
/gene
```



```

/translation="MKMSVLSLSFSPGHAPQRLTIKRVREYLQKGLQDAVVADQIIP
RSFSVLSMR"
gene
5354. .5677
/locus_tag="BR1026"
CDS
5354. -5677
/locus_tag="BR1026"
/notes="similar to GP:15156412; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ferredoxin, 2Fe-2S"
/protein_id="AAN29948.1"
/db_xref="GI:23347848"
/translation="MTKIVFVSADGATRTEADNGSSVMEAAIRNGIPGIDAECCGA
CACATCRVYVDDWADTVGGDPDMEEDMLDFAYEVRPTSRLSQCIKRVTDLLEGLVIVQ
PERQN"
gene
5714. .6316
/locus_tag="BR1027"
/notes="this region contains an authentic point mutation,
causing a premature stop, and is not the result of a
sequencing artifact; thioredoxin reductase, putative,
authentic point mutation"
/complement(6321..6941)
/locus_tag="BR1028"
/complement(6321..6941)
/locus_tag="BR1028"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN29949.1"
/db_xref="GI:23347849"
/translation="MNTKKVLAALAAAMAILPQAAAAIIFRKFOYTYINGTKTAADIL
XALSRSPFLAKTQGHGPGAAERFQAKVYRGREPKCKVDVYVNVHAKVSLPRWK
QRRKASPLALWNTLODTRHEESHVIARSHASEMERISLRSLRSCADSLRADI
DKTVSRVMEADHEAQYDFRVETINFGNRFLLTYGLERMKQTGH"
gene
7099. .7938
/gene="folp"
/locus_tag="BR1029"
7099. -7938
/gene="folp"
/locus_tag="BR1029"
/notes="similar to SP:P28822; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydropteroate synthase"
/protein_id="AAN29950.1"
/db_xref="GI:23347850"
/translation="MTKXEQAHRSLSHLGKTAIINGVLNVTSPFSQGGRYNOLDRA
IAAETMLEEGATIVDVGSETPGATAVDAQTEARVVFIRELVRFDCIISVDYI
RASTLARVEAGAHVINDWGLQPEPEIAQTEARGAGVIMHTSRDLTNPVDVIEDQ
FNFILNRSLKAGADIGSPRIVLPDGFQKDEDEIDALMARVSELOAFGFPILVGTG
RKRFVGMGTQGDPPNRDITGAATSVLRLAGADIIFRVHNVAFNRDALAVADAILQNN
VRI"
gene
7957. .8316
/gene="folB"
/locus_tag="BR1030"
7957. .8316
/gene="folB"
/locus_tag="BR1030"
/notes="similar to GB:S45214, GB:Z12962, GB:D28462,
SP:P28751, and P10:36136; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydroneopterin aldolase"
/protein_id="AAN29951.1"
/db_xref="GI:23347851"
/translation="MYTIRIMNCAFFAHGHGVFDEEHKLGQRFYDAVLDDVEGNSLR
DNIDGTVHVGIAFVSPIRIMIGRQRYLIETLALOKALVALVFPQIKRAEITVRKPNR
PVSGLDYDEVTVVYYPQ"
gene
8313. .8846

```

CDS	/gene="folk"
	/locus_tag="BR1031"
	8313..8846
	/gene="folk"
	/locus_tag="BR1031"
	/note="similar to GP:15074743; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="2-amino-4-hydroxy-6- hydroxymethylidihydropyridine pyrophosphokinase"
	/protein_id="AAN29952.1"
	/db_xref="GI:23347852"
	/translation="MSTPPKPLAWLGLGNNIDPIISMASRLMDGRSDTRVVSUS PVYRPPKGKDQAFWENACQALETALQPLELIAACLVERSILKRVRLERMGPRIIDI DILVMQDETGRAITMQSPAELPHPRHERAFVLVELNDIAPSLPVGARTVAEWAADC DRAEMKARTDAGWALE"
gene	complement(8961)..9518)
	/locus_tag="BR1032"
CDS	complement(8961)..9518) /locus_tag="BR1032" /note="similar to GP:14021684; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAN29953.1"
	/db_xref="GI:23347853"
	/translation="WDSETELMKNCTAACLTLLPVLSLSVSPMAADKDKOFFQTIEG QNSGGEIVAGKYGTFRVCNLAGTTFDDTVGMTLOGTCRVGVFSOPMKATVTRYGDG YAKNGDGADGGDLVDVSGVNGKRVFGLNRQLNGLMARVSDENTWNVTVSRYRE KELVPVIGMSLRVDVTAVGSIAKN"
gene	complement(9770)..10843)
	/locus_tag="BR1033"
CDS	complement(9770)..10843) /locus_tag="BR1033" /note="identified by match to TIGR protein family HMM TIGR01620"
	/codon_start=1
	Query Match 34.9%; Score 115.2; DB 1; Length 12376; Best Local Similarity 60.6%; Pred. No. 3.1e-13; Matches 189; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
Qy	19 GTGATCTTCGGCAGCGCGGCTTCGAGCATCTGTCGAAACCGAAGCGGGAGTATCG 78
Dd	5366 GTATTCTGATCCGCCGATGGCCCCACCGGAACCGAGGTCGAGGCCACAACGGGTCCAGC 5425
Qy	79 CTGATGAGAGCGCGCTTCTGAACCGCGTGACGGTATCGAAGCGGTTTCGGGGGCGCC 138
Dd	5426 GTTATGGAACGGCCATTGCBAACGGCAATCCCGGTATTGATGCGGAATGGGTGGCGCT 5485
Qy	139 TGTCCTCGGCCACGTCGCCACTTTACGTGGCCCGGAGTGGCTAGATGGCTGGAACCG 198
Dd	5486 TGGCGCTTCGCAACTCGCCATGCTATGTCGATCACGACTGGGCCGATACGGTCGGCGGG 5545
Qy	199 CCGAGTCAGACCGAAGACGAATGCTCGATTGCGTAGCGGAACGTGCGCGCGCAATCGCGG 258
Dd	5546 CCGGATCCAATGGAAGAGATATGCTGGATTTCGGCTATGAAGTCGCCCGCACTCGCGC 5605
Qy	259 CTGTCCTGCCAGATCCGCCCTTACCAGCTGCTCGACGCCCTGACCCCTGGAACTGCCGAAG 318
Dd	5606 CTTTCCTGCCAGATCCGCGTTACCGACGATCTGGAGGGGCTTGTCGTTCAAGTCCCAGAA 5665
Qy	319 GCACATCATGA 330
Dd	5666 CGCCAGAAGTGA 5677
RESULT 4	
AE009536/c	AE009536 linear BCT 20-MAR-2003
LOCUS	12429 bp DNA
DEFINITION	Brucella melitensis 16M chromosome I, section 93 of 195 of the

[illegible]

```
/translation="MSQGIKMTIKIVFVSADGATRTVEADSSGVMEAAIRNGIPGID
AECGGACACATCHYVYDDADITVGDPDPMEEDMLDPAYEVPTSLSCQIRVTDLE
GLVQVPERQN"
/gene="BMEI0960"
/complement(4398. .4679)
/codon_start=1
/transl_table=11
/product="Hypothetical Protein"
/protein_id="AAU52141.1"
/db_xref="GI:17982917"
/translation="MWSKSVLSFISPGHAFQRLTIKVKREYLQKXHLQDAVVADQDS
PVFFISSYKIIIRFSPCGKVLGAPAFADFDAGAWRSRGAAYGG"
/gene="BMEI0961"
/587. .9567
/gene="BMEI0961"
/587. .9567
/gene="BMEI0961"
/codon_start=1
/transl_table=11
/product="Kinesin-like protein"
/protein_id="AAU52142.1"
/db_xref="GI:17982918"
/translation="MFWGFAQLVKRAQEMONARSMTDAALKLLOPEAVAGRVSTLG
QAVREVAANNEGIERLARAVELETLVQSEVQLERAYSQDNVRSLSVSDLGNERE
AVVSHAERKRSISGAHEQREELUSSNIIIRNVLSQALSGALLDGERLISIN
ESGAIDAIEKRTDIGNRTIGEPANLIDTRIATDEQRTSGKLTQALDERT
TGIANVLQATKTVWSEFDRLANLENTSDRGSLLAEFAEFAHALDSTENKNAAL
EATROQINENLIATRIAEFTSGRNSLGMVDDIKVIGDDLTGISVGVNLIDK
AAPEGLASRDLLAATLEGETRVSIVGHADTLAARTDIOAIDASASALNEV
TDKHAAILTERTSALOQMSNVSLDASPADHARSLEERTASLHNVIAGNOAMLAQL
MDERAAAMRENFRNVRVARTFDERIGSLQSLIAENQSALEILDERAARMAESLGA
GRDAFETALGEHNRVBERVSGLOAVLNNHDAHVDHEQETIERAAEIRSLTLD
STASLTLDHGBILSRTSLQNAIANSAAISSAFEGGTGIIERTETMBKALSI
GVNVRALQSQAVAKREKTEGEAAATLAAAKAGALDGFESFASRLIDNLS
GTEARLGERDAIAGRLGDIERSLTDIGTIEARIADTAARTSETLASHSESPASVA
DLNAGTRELIVERADATRLDIDIGSRTIWEIGSVIARIQVDTTAVTLEERTRELN
AVLAARSOEITKIINDAEPLVOLFADSLGSLAQCLQEATHTAATDRLSERNAALNAL
ASTRAETIAVQQAQKVSNDVSELIDRLAASGELGKLIDATRLNADIDGRVDSST
TSFVNTNRAQMFQASTGLDLSNIGLITSLNSLTQIADIADRDDEHKGVLSSASD
MINEAQNGLVSTIEERQALDKLASLIVERSEKVEKLMQSFALVASAFQADMPAEKE
SAEKRVESVDIEVQAQKFSAADEIRRTASEIRABELDTTRELKRGVLDMPAEAKE
TTTAMRKAVSELQINALELAIIVNKSRLVDVGEGRADRPVSPAPASRPAPQVYRA
PIALACTDVAVRPVPEPNOKPAPASAEKPRGWSDDLARASEREEBAAPKVPQ
AATPRSPHVSELSNLSVDIARALDHEASVELMDRVYRGERNVFRRLYTLKGGQTF
DDIKRKVQDGEFRADVDRVNDPFRLELLEVDVANDRDNDVNTQIYLSIDTGKVTMLAH
ASGRIR"
/complement(9807. .10364)
/gene="BMEI0962"
/complement(9807. .10364)
/gene="BMEI0962"
/codon_start=1

Query Match 34.9%; Score 115.2; DB 1; Length 12429;
Best Local Similarity 60.6%; Pred. No. 3.1e-13;
Matches 189; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 19 GTGATCTTCCCGCAGCGCGGGCTTCGAGCATCTGTCGAAACCGGAGCGGAGTATCG 78
Db 4267 GTATTCGTATCCGCGCATGGCGCCACGCGACCGAGGTTCAGCGCCGACGGGTCCAGC 4208
Qy 79 CTCATGGAAGCGCGCGTTCGAAACCGCGTGAACCGGTATCGAAGCGGTTTCGGGGCGCC 138
Db 4207 GTTATGGAAGCGCGCATTCGCAACCGCATTCGCGGTATTCGATGCGCAATCGCGTGGCGCT 4148
Qy 139 TGTGCTCGCCACGTCGCGTTCGTTAGTGGCCCCGAGTGGCTAGATGCGCTGAACCG 198
Db 4147 TGGCCTCGCAACCTGCGCATGTCGTATGTCATGACGACGTGGGCGCGATACGGTCGGCGG 4088
Qy 199 CCGAGTGAGACCGGAGACGAATGCTCGATTGCGTAGCGGAACGTCGCGCGCATTCGCGG 258
Db 4087 CCGGATCCAAATGGAAGAGGATATGCTGATTGCGCTTATGAAGTCCCGCCGACCTCGCGC 4028
```

```
Qy 259 CTGCTCTCCAGATCCGCTTACCGACCTGCTCGACGCGCTGACCTGGAACCTGCCGAAG 318
Db 4027 CTTTCTCTCCAGATCCGCTTACCGACCTGCGAGGCGCTTGTCTTACGTCGCCGAA 3968

Qy 319 GCACAGTCATGA 330
Db 3967 CGCCAGAACTGA 3956

RESULT 5
AP002995/c 346897 bp DNA linear BCT 15-MAY-2001
LOCUS Mesorhizobium loti DNA, complete genome, section 2/21.
DEFINITION AP002995 BA000012
ACCESSION AP002995.2 GI:14021442
VERSION 1
KEYWORDS Mesorhizobium loti
SOURCE Mesorhizobium loti
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE 1 (sites)
AUTHORS Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
Watanabe,A., Idegawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
Takeuchi,C., Yamada,M. and Tabata,S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
JOURNAL MEDLINE
PUBMED 21082930
REFERENCE 2 (bases 1 to 346897)
AUTHORS Kaneko,T.
Direct Submission
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp, rhizobase/,
URL:http://www.kazusa.or.jp/rhizobase/,
Tel:81-438-52-3935 (ex.2338), Fax:81-438-52-3934)
On May 11, 2001 this sequence version replaced gi:11994963.
FEATURES
source
location/Qualifiers
1. 346897
/organism="Mesorhizobium loti"
/mol_type="genomic DNA"
/strain="MAFF303099"
/db_xref="taxon:381"
74. 160
/note="trns-TGA"
complement(240. .1109)
/gene="msl18587"
complement(240. .1109)
/gene="msl18587"
/note="unknown protein"
/codon_start=1
/transl_table=11
/db_xref="GI:14021443"
/db_xref="BAB48055.1"
translation="MYDOTTTGITATPVTYVIRAGDVANPOAVHKGTFYHFPREG
AGNVLIVSNKHVLVGKSWLEDFALLDNQKRYFGPANTRELTKNQLPIBEHPDPTV
DLALPIPTIVAMRGQKGLFNLGLRETTAFDNQIALRLAASVLMVGFPSGIMDE
KNLPLVVRGILAVNYLTDYGETNFVVDIAAFGSGSPVFAFFENMLPDGSGVTM
LAQGVYFLVHLHSGPSMTAKIYVPVPTSDQTAQTMMLHLGYCYKASRVEBELVE
TIRAKISPTASG"
complement(1118. .1657)
/gene="ml10476"
complement(1118. .1657)
/gene="ml10476"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48056.1"
/db_xref="GI:14021444"

gene
CDS
```



```

NGEAILVNLILMSLFAIHSVMARPAFRWAKFLPLACORSTYVLLSSILLILFLW
QMRPTPTVMTSGTGAIAWLLIGVHWLWLIAPASTHMDHFDLFLGLQAFAVFRGTI
SQSFRTPYKIVIRHPLMLGLFLAFWATPANTAGHLLFALANTAYILVALQFEERDL
IAEFGATQDQYRRRVRVPLFAFRRTTDRKSPRPVCAPR"
/ gene
5192 .6169
/ gene="bl15742"
/ note="ORF ID:bl15742
putative quinone oxidoreductase (EC 1.6.5.5)"
/ codon_start=1
/ transl_table=11
/ protein_id="BAC51007.1"
/ db_xref="GI:27354019"
/ translation="MKAVLCSSFTGPADLRLLSEIDPRPAPGDEILIDVHAASVFMQO
LMVSLYQMRPTTPVPTGTEASGIVVAVGDKVTAPAPGRVACSSWTGGVABRMTAKE
SKSVRLPGVAFESADTALVHNHNTAYVALVERARAQAGETVITCAAGGVGLAAVDLG
RHFGLRVAGVGSDDKAALVYGYGASEIVNRSSEDLRIKSIITSGGQIDVGFNVGG
AVFEQWALMKWGGKLLPIGPASGEIPOVPMNLPLLNYSIIIGVFVGAWEAFESAEEA
OMNDTLMOLLADQIRPHIDCVLPLEBAGDGWEAVANETVQGRIVLKIR"
6202. .6954
/ gene="bl15743"
6202. .6954
/ gene="bl15743"
/ note="ORF ID:bl15743
hypothetical protein"
/ codon_start=1
/ transl_table=11
/ protein_id="BAC51008.1"
/ db_xref="GI:27354020"
/ translation="MTWIIIDKRYCYPNNGGVYCGRLARHIFGAAEVTLRPPPL
DRPLDVAITDDGTDLRGAKLVATGRPTNVLELRLAKESPEEARSABLTPVPKPHH
PLTFCVCPAPAKGDKLRI FAGLRDVRNVAALASWIPDNLAAEDGLVASEFLWS
ALDCTGTACYNQSDSFDQAPLLGRMSARIEARPRPGRCVITAWPTGRDGRKRT
ABAAVDETGLLAVARATWIAVEREVQLGCG"
7289. .8209
/ gene="bl15744"
7289. .8209
/ gene="bl15744"
/ note="ORF ID:bl15744
unknown protein"
/ codon_start=1
/ transl_table=11
/ protein_id="BAC51009.1"
/ db_xref="GI:27354021"
/ translation="MRLPAPALLFTSLFGFLLCDSAPANPTVCPM5AALEDDT
LYAFQRTWTPVISASEPEAPQKRWITLYFVVKTAGDDRAGAIVIKTARLQPSLTSR
PYQVRLRSGVGNCSQSYLSVSTLYAVEAYHDLGYDRGEILERLPGSDTGLTT
LOHRTGYDATYASRDQKCRDDBKRYDGSYEARSNSQSFVTDIVDHGIGSAAS
QAASGGRVWASLIYPPASANRENLRERVEIKPYQTWAGTACTVPRISPRGTAQVIR
VNDLRKQVDFGRKXIRLGPITGGEF"
/ complement(8231. .10771)
/ gene="bl15745"
/ complement(8231. .10771)
/ gene="bl15745"
/ note="ORF ID:bl15745
unknown protein"
/ codon_start=1
/ transl_table=11
/ protein_id="BAC51010.1"
/ db_xref="GI:27354022"
/ translation="MGAPVLFKRSILTOAGATAIARQHEAAVLQDERAAVAVGRKRAV
TGPWIGLGGIRSATFCLGALORFANSGLLKHFDDYTSVSGGVAASAVOMKWH
DPTDAGANFPVGSQTSQSLASENSLKYLRWHSFLAPGGGIS"MSGIAYLLEFLI
SVFFWPLAIAFWLVFVATGHAFYESYGTWGLTVPKLEANTLSLFWIGICUSAF
GVTFIAYILLALIGTIFLAIISILIPPTENGRRIARAGCAAVGAGLLVSY
FTFHGAFDFDLPATDVPTSSALLIATAGFLGACAIIVGILQSLGILEFGYAVRR
KYLEAATWPTLIACLTGSLFPFVEIADSTKYSTGSKALIGITFVASGFLSGLY
GHFVQARLPALATVAATLAAGAFLLIIVIYAGARFLIQQHLLADQAQYCREF
FVALVFSVAFGWSNVNVLGHRFVRDRMEAFMTTBOKAADPNYSDDRLPLAA
IIMPKOTSPLVPLINTNVLLNDQDVALRGGSFLLSPCYVGSFTWMLTSDHG
KXHGFWTLASAAASAAANAGFTGCTITDRLLISTVMMLFNRLRLGLMGAPQASR
GPNYFSPGLKTVFGRSGTQSDSDFLELSDGGHFNGLGAYELIRRCDDVIVLIDSEED

```

```

PTAMSAALASVCQVRVEDFVITKIEAEADTIAPAGDMGYPAGAKFAKKSIFYASIDY
PAEHRGTIADQARTGTFFVYKSGMIKQISFTKAKYKYPDFPNOSTVDFQFDSAQFE
AVRDLGWSVENMASAFGLDAATAADDFVRLITGTAPPPAAPAPAPAG"
/ complement(10900. .12096)
/ gene="bl15746"
Query Match 34.5%; Score 114; DB 1; Length 300140;
Best Local Similarity 62.8%; Pred. No. 3.9e-13;
Matches 177; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
QY 49 CATCTGTCGAAACCGGAGGATCGCTCATGGAAGCGGCCCTTCTGTAACGCGGTG 108
Db 164676 CGTACGGTGGAAATCGAGAACGCGCAACCGTGATGAGCCGCATCCGACACAT 164617
QY 109 GACGGTATCGAACCCGTTTTCGGGGGGCGCTGTGCTCGCCACGTCGCCACGTTTACGTT 168
Db 164616 CCGGGATCGAGCCGGAATCGCGCGCGCATGTGCTCGCGCACCTGCTGTATGTC 164557
QY 169 GGGCCCGAGTGGCTAGATCGCTGAAACCGCGCGAGTGAGACCGAAGCAATGCTCGAT 228
Db 164556 GACGAGACCTGGCGTGAAGAGGTGCGGACGCGCGCATGGAAGAGCATGCTCGAC 164497
QY 229 TGCCTAGCGGAACGTGCGCGCATTCGCGGCTGTCTGCGAGATCCGCTTACCGACCTG 288
Db 164496 TTCGGTTTCACGTGCGCGCGCAACTCGCGCTGTCTGCGAGATCAAGGTCTCCGACGAG 164437
QY 289 CTCGACGGCTGACCTCGGAACTGCGGAGGACAGTATGA 330
Db 164436 CTCGACGGCTGTGCTGTGTCGACGCGCGGACGCGCGCTGA 164395
RESULT 7
LOCUS CCDFXB 1351 bp DNA linear BCT 16-APR-1992
DEFINITION C.crescentus DNA for 16S rRNA and fdxX gene.
ACCESSION X51607
VERSION X51607.1 GI:40414
KEYWORDS 16S ribosomal RNA; fdxX gene.
SOURCE Caulobacter vibrioides (Caulobacter crescentus)
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
Caulobacteraceae; Caulobacter.
REFERENCE 1 (bases 1 to 1351)
AUTHORS Anemiyi,K.
TITLE Conserved sequence elements upstream and downstream from the
transcription initiation site of the Caulobacter crescentus rRNA
gene cluster
J. Mol. Biol. 210 (2), 245-254 (1989)
90096149
MEDLINE 2600967
PUBMED
REFERENCE 2 (bases 1 to 1351)
AUTHORS Anemiyi,K.
TITLE Direct Submission
JOURNAL Submitted (02-APR-1992) K. Anemiyi, National Inst of Health, Lab of
Viral and Mol Pathogenesis, Bethesda MD 20892, USA
FEATURES
Location/Qualifiers
1..1351
/organism="Caulobacter vibrioides"
/mol_type="genomic DNA"
/strain="CB15"
/db_xref="taxon:155892"
101..421
/ gene="fdxX"
101..421
/ gene="fdxX"
101..421
/ codon_start=1
/ transl_table=11
/ protein_id="CAA35950.1"
/ db_xref="GI:40415"
/ db_xref="GOA:P37098"
/ db_xref="SWISS-PROT:P37098"
/ translation="MAKITYIQHDGAEQVLDKVEGLTMEGAVKNNVPGIDADCGGAC
ACATCHVYDVAWLDTGDKSANEESMLDPAENVBPNSRLSCQIKYSDALDGLVRLP

```

ESQH"	543..>1351	/product="16S ribosomal RNA 5'end"
organism="Caulobacter crescentus CB15"		
mol_type="genomic DNA"		
strain="CB15"		
db_xref="taxon:190650"		
complement(179..254)		
gene="trRNA-Ala-2"		
complement(179..254)		
gene="trRNA-Ala-2"		
product="trRNA-Ala"		
complement(322..398)		
gene="trRNA-11e-2"		
complement(322..398)		
gene="trRNA-11e-2"		
product="trRNA-11e"		
complement(589..2027)		
gene="CCrRNA16S"		
complement(589..2027)		
gene="CCrRNA16S"		
product="ribosomal RNA"		
complement(2688..2909)		
gene="CC3523"		
complement(2688..2909)		
gene="CC3523"		
note="identified by Glimmer2; putative"		
codon_start=1		
transl_table=11		
product="hypothetical protein"		
protein_id="AAK25485.1"		
db_xref="GI:1342525"		
/translation="MTGARLGLGLHLLTSKERRYIVARGLPVAGWPTSPRPE		
RHRDTRPPKSGICPDASLETALOSTQ"		
complement(2912..3232)		
gene="CC3524"		
complement(2912..3232)		
gene="CC3524"		
note="identified by match to protein family HMM"		
codon_start=1		
transl_table=11		
product="ferredoxin, 2Fe-2S"		
protein_id="AAK25486.1"		
db_xref="GI:1342525"		
/translation="MAKITVQHDGAEQVIVKPLTMEGAVKNNVPGIDACGAC		
ACATCHVYVDEAWLDTGDKSAMESMLDPAENVEPNRLSCQIKVSDALDGLVLRP		
ESQH"		
complement(3276..4514)		
gene="CC3525"		
complement(3276..4514)		
gene="CC3525"		
note="identified by match to protein family HMM"		
codon_start=1		
transl_table=11		
product="ferredoxin reductase"		
protein_id="AAK25487.1"		
db_xref="GI:1342525"		
/translation="MSAEVNONACVIVGAGHAGGSVAALROYHGHGRIVLIGDPL		
LPYQRPFLSKWLGADADSLKPGAWYADNNMLRGVVAERINRSKTVLALASG		
EVIPYDFLVLTAGARDELPTFGADLAVLATAADAEILKNALGPKDLAVVGGY		
VGLEAASARALGSHAVVIERESVLARVACETLSHFQDYHGHKGHVAFELNAGVA		
EGHDGHTVGRFNDGRVVAACDVALGVAVNDELAKDAGLSTANGVVDLEARTDDP		
SIFATGDYTHRELPLYDRQPELESVPNALQQAASAILGRPCPAPETFWFMSDQV		
LKIQAGLQFPDADQVVRGDAAKFAVFLKGLVQAVEAVNPPEFWMAGKQLIAK		
TPVDANKLADFSVSKVEVA"		
complement(4643..5425)		
gene="CC3526"		
complement(4643..5425)		
gene="CC3526"		
note="identified by match to protein family HMM"		
codon_start=1		
transl_table=11		
product="succinate dehydrogenase, iron-sulfur protein"		
protein_id="AAK25488.1"		
db_xref="GI:1342525"		
/translation="WVELALPKNSQVSKGHWPAAGAKVKVTKIYRYDPESGNP		
ESQH"		
complement(3276..4514)		
gene="CC3525"		
complement(3276..4514)		
gene="CC3525"		
note="identified by match to protein family HMM"		
codon_start=1		
transl_table=11		
product="ferredoxin reductase"		
protein_id="AAK25487.1"		
db_xref="GI:1342525"		
/translation="MSAEVNONACVIVGAGHAGGSVAALROYHGHGRIVLIGDPL		
LPYQRPFLSKWLGADADSLKPGAWYADNNMLRGVVAERINRSKTVLALASG		
EVIPYDFLVLTAGARDELPTFGADLAVLATAADAEILKNALGPKDLAVVGGY		
VGLEAASARALGSHAVVIERESVLARVACETLSHFQDYHGHKGHVAFELNAGVA		
EGHDGHTVGRFNDGRVVAACDVALGVAVNDELAKDAGLSTANGVVDLEARTDDP		
SIFATGDYTHRELPLYDRQPELESVPNALQQAASAILGRPCPAPETFWFMSDQV		
LKIQAGLQFPDADQVVRGDAAKFAVFLKGLVQAVEAVNPPEFWMAGKQLIAK		
TPVDANKLADFSVSKVEVA"		
complement(4643..5425)		
gene="CC3526"		
complement(4643..5425)		
gene="CC3526"		
note="identified by match to protein family HMM"		
codon_start=1		
transl_table=11		
product="succinate dehydrogenase, iron-sulfur protein"		
protein_id="AAK25488.1"		
db_xref="GI:1342525"		
/translation="WVELALPKNSQVSKGHWPAAGAKVKVTKIYRYDPESGNP		
ESQH"		
complement(3276..4514)		
gene="CC3525"		
complement(3276..4514)		
gene="CC3525"		
note="identified by match to protein family HMM"		
codon_start=1		
transl_table=11		
product="ferredoxin reductase"		
protein_id="AAK25487.1"		
db_xref="GI:1342525"		
/translation="MSAEVNONACVIVGAGHAGGSVAALROYHGHGRIVLIGDPL		
LPYQRPFLSKWLGADADSLKPGAWYADNNMLRGVVAERINRSKTVLALASG		
EVIPYDFLVLTAGARDELPTFGADLAVLATAADAEILKNALGPKDLAVVGGY		
VGLEAASARALGSHAVVIERESVLARVACETLSHFQDYHGHKGHVAFELNAGVA		
EGHDGHTVGRFNDGRVVAACDVALGVAVNDELAKDAGLSTANGVVDLEARTDDP		
SIFATGDYTHRELPLYDRQPELESVPNALQQAASAILGRPCPAPETFWFMSDQV		
LKIQAGLQFPDADQVVRGDAAKFAVFLKGLVQAVEAVNPPEFWMAGKQLIAK		
TPVDANKLADFSVSKVEVA"		
complement(4643..5425)		
gene="CC3526"		
complement(4643..5425)		
gene="CC3526"		
note="identified by match to protein family HMM"		
codon_start=1		
transl_table=11		
product="succinate dehydrogenase, iron-sulfur protein"		
protein_id="AAK25488.1"		
db_xref="GI:1342525"		
/translation="WVELALPKNSQVSKGHWPAAGAKVKVTKIYRYDPESGNP		
ESQH"		
complement(3276..4514)		


```
/organism="Rhodobacter capsulatus"
/mol_type="genomic DNA"
/strain="B10"
/db_xref="taxon:1061"
379..1761
/gene="ORF E0"
379..1761
/gene="ORF E0"
/codon_start=1
/transl_table=11
/protein_id="CAA72164.1"
/db_xref="GI:2094850"
/db_xref="GOA:O06439"
/db_xref="SPTREMBL:O06439"
/translation="MFPSLSIRFADLVVQKPAVVMITTTVSPPAGMPGPPGGS
PFEFRFRFEMNMQGPGOPORHDAALGSGFVLSADGKIVTNHVIKGGDGNVVADE
ISTEPANGKLDKIVGTDPKTDIALKVKSEKPLPVSFSGDSTREGDVMAMGNP
LGQFSIVSGIVSARNRELSGTVDCCIOTDAALNRNGSGPLFSMDGKIVGNTALIS
PNGSICIGFSMASNVKAGQLOFEGFATRGVGLVKIQDLPDPAEAMSLATPEGA
MVSEVPGPAKAVGMSGDV11SPAGAPVDTKSLRRVADAFVGAADVVMRAGKP
VTQVTVGQRELAESEAAAGTAPALAGAVLGMVLSPTIDAMRAERLNPDLTGLM
VTDIDTGAAPDKLGRADVIVBAGQQPVTSLLDLKARVAEAKGCKSVLLIRREG
EPRFVAVPLO"
2540..2863
/gene="fdx"
2540..2863
/gene="fdx"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="CAA72162.1"
/db_xref="GI:2094851"
/db_xref="GOA:O09767"
/db_xref="SPTREMBL:O09767"
/translation="MAKIIETEHNGTRHEVEAKPGLTVMEARNDRNGVPGIDACGGAC
ACSTHAYVDPAPWVKLPKALPTETMDIDFAYEPNPATSLTQIKVTSLLDLGLVHL
PEKQI"
3251..4075
/gene="ORF E1"
3251..4075
/gene="ORF E1"
/codon_start=1
/transl_table=11
/product="PurU-like protein"
/protein_id="CAA72163.1"
/db_xref="GI:2094852"
/db_xref="GOA:O06440"
/db_xref="SPTREMBL:O06440"
/translation="MTTYVLTVNCSTRGIIVAAITANYLADSGCNLTDSNYDDLTLGQ
PFMRVTFVQGTALDSLKAGFEPVQVDFMWAHDAEKVKVLLMVSFNGHCLANDL
LYNRIGALPVEIVGVVSNHMTYQKVVVNHDIIPPHIKVTKENKPEGEGLLDVVEES
GGELVVLAVGVIQDKICOKMSGKIKIHSFLARFKGNPYKQVVERGVKLVGTVSH
YVITALARTDHRAGNPHHPKAPRGLRQPRFCRGAGARFGDPRFCAAPRLHQ"

Query Match 32.2%; Score 106.4; DB 1; Length 4151;
Best Local Similarity 61.4%; Pred. No. 1.9e-11;
Matches 189; Conservative 0; Mismatches 116; Indels 3; Gaps 1;

QY 20 TGATCTTCGGCGCGCGCGCTTCGAGCATCTGTCGAACCGAGCGGAGTATCGC 79
DB 2550 TCATCTTCATGAACACACACCGCACCCCTCGTCAGAGGTGGAGCGAAACCGCGGCTGACGG 2609

QY 80 TCATGGAAGCGCGCCTTCTGAACCGCGTGGACGGTATCGAAGCGGCTTTTCGGCGGCGCCT 139
DB 2610 TGATGGAAGCGCGCGCACATGCGGTGCGGGCATCGACGCCGATTTCGGCGCGCCT 2669

QY 140 GTGCTTCGGCCACGTCACGCTTACGTTTGGCCCGCGAGTGCCTAGATCGCTGAACCGC 199
DB 2670 GCGCTGTTCGACCTGCCACGCTATGTCGATTCGGGCTCGGTGACAAAGTCGCGAAG 2729

QY 200 CGAGTGAGACCGAAGACGAATGCTCGATTGCGTAGCGGAAC---GTGCGCGCGCATTCGC 256
```

Db 2730 CACTGCGCAGCAGACCGACATGATCGATTTCGCTATGAGCGCAACCCCGACCTCGC 2789

QY 257 GCCTGCTCTGCAGATCCGCTTACCGACCTGCTCGACGGCCTGACCTGCAACTGCCGA 316

Db 2790 GCGTGACCTGCAGATCAAGGTCACTGCTGCTGACGGGCTGCTGCTGATCTGCGCG 2849

QY 317 AGGCACAG 324

Db 2850 AAAAAACAG 2857

RESULT 10

AB121977

LOCUS

DEFINITION

AB121977 6407 bp DNA linear BCT 11-OCT-2003

Xanthobacter polyaromaticivorans dbd gene cluster (dbdA, dbdB,

dbdCa, dbdCb, dbdCc, dbdD, dbdE), complete and partial cds.

AB121977

AB121977.1 GI:37651307

VERSION

KEYWORDS

SOURCE

ORGANISM

Xanthobacter polyaromaticivorans

Xanthobacter polyaromaticivorans

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

Hyphomicrobiaceae; Xanthobacter.

REFERENCE

1

AUTHORS

Hirano,S., Haruki,M., Tkano,K., Imanaka,T., Morikawa,M. and

Kanaya,S.

TITLE

Gene cloning and characterization of a unique dibenzothiophene

dioxygenase from Xanthobacter polyaromaticivorans strain 127W

REFERENCE

2 (bases 1 to 6407)

JOURNAL

Unpublished

AUTHORS

Morikawa,M., Hirano,S., Imanaka,T. and Kanaya,S.

TITLE

Direct Submission

JOURNAL

Submitted (07-OCT-2003) Masaaki Morikawa, Osaka University,

Material and Life Science, Engineering; 2-1 Yamadaoka, Suita, Osaka

565-0871, Japan (E-mail:morikawa@mls.eng.osaka-u.ac.jp,

Tel:81-6-6879-7443, Fax:81-6-6879-7443)

FEATURES

Location/Qualifiers

1..6407

/organism="Xanthobacter polyaromaticivorans"

/mol_type="genomic DNA"

/strain="127W"

/db_xref="taxon:225625"

417..740

/gene="dbdA"

417..740

/gene="dbdA"

/codon_start=1

/transl_table=11

/product="ferredoxin"

/protein_id="BAC98954.1"

/db_xref="GI:37651308"

/translation="MSIKVTIEHDGIQRTVEAHPGASIMEIAVNVGPGIDAECCGA

CACATCHIYVDAWLKCLPAPSDSEAMLGFAAHRQTNLSLSCQLRLQEHGDIVVT

PAFQ"

814..1866

/gene="dbdB"

814..1866

/gene="dbdB"

/codon_start=1

/transl_table=11

/product="gentisate 1,2-dioxygenase"

/protein_id="BAC98955.1"

/db_xref="GI:37651309"

1975..3276

/gene="dbdCa"

1975..3276

gene

CDS

gene

CDS

```
gene
/feature="dbdCa"
/codon_start=1
/transl_table=11
/product="alpha subunit of terminal oxygenase"
/protein_id="BAC98956.1"
/db_xref="GI:37651310"
/translation="MTLAPHTPSGKYNAREIRALVNTDITGRYSPEIYTSBDLYOLEL
BIFPARGVMCHETQIQKAGDYOAAYIGEDPVVVRKODNSIRVFLNOCRRHGRWIC
RVDGNAKATCPYHGAYDITAGNLVSVPMEQEARGGKLNKAEMGPKOARVETVYGLI
FANWESAPNDITYLDGAKFYNDIMLDRAGEGTEVITGVQKFIIPANWKFAAEFGSD
AYHAGTITGSGIAGVPAQDITSLRKVPPTDGMNVIYIGNHGGCGLFSRPNPFIYVICG
POLTYLTGSPAYEBAVRLGGADVNLCHNVPFNLSTGLNTVVMQPRGNEM
ELMTFTLVDRTPDNKKEEMRNMYRTFASGVFEQDDGDNNDIQHVLRGHVARQOP
FNIEMAQHTVQSTSIAPGPPGESPAYTVAEESGTQHLQALG"
/3401..3964
/gene="dbdCb"
/3401..3964
/3401..3964
/gene="dbdCb"
/codon_start=1
/transl_table=11
/product="beta subunit of terminal oxygenase"
/protein_id="BAC98957.1"
/db_xref="GI:37651311"
/translation="MLDTSASQAFQVKPELISVELQHEIQFYVWEAKLMADREWEH
FSLMKDIXYAPLPSRSHAEAKLEYTDKGFHPADNWSLDRGRKITSYVGS
ENPARKLRHIVGNVLWIDRQADYVDVSTSLTYNROERQTDFFAAERQDIRRVD
ACGFIVERKILIDQSTLLSNNISFF"
/3979..4311
/gene="dbdCc"
/3979..4311
/3979..4311
/gene="dbdCc"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="BAC98958.1"
/db_xref="GI:37651312"
/translation="WPNVYVLPDLPLPGDKMVFNGAEPILICNVQGVYAVQDTC
HTWSLCDYLDGHIVECSLHNAKFDVDTGEVKALPACKALKIPKIKENGEIYVDE
RQTSBGGT"
/4312..5139
/gene="dbdD"
/4312..5139
/gene="dbdD"
/codon_start=1
/transl_table=11
/product="dihydrodiol dehydrogenase"
/protein_id="BAC98959.1"
/db_xref="GI:37651313"
/translation="MCKNPFVLTGSGSLGKAIVERFLAEGAKVAILDRSEERVQEV
VAMFGGVVGTSGDVRSMADNKAARVCVEAFKIDITLVGNAGVWDYNASIAGTSEEK
LABAIDELVGINLKGILYLAAKAALPELYKGNALFTVSNAGFPYGGGVLYTAKKHG
VVMGKOLAHAWAPHIRVNGVAPGGIGISRGFPSSLQEGIRFSLPLDDLMKWLPL
EKAFYASEVTGCVYFPAERKDNPGATGSLVNFDDGIGMKGFMSPNMGKELDEKFGA"
/5233..6407
/gene="dbdE"
/5233..6407
/gene="dbdE"
/codon_start=1
/transl_table=11
/product="salicylaldehyde dehydrogenase"
/protein_id="BAC98960.1"
/db_xref="GI:37651314"
/translation="MSELLRVGDEGSAADCAVLERSPADGVVSVRAAKRADVE
AAIATAHAYIEMSKTSPSHRMLLKADELEKRADESLVEAMKDEVGSAELWARFNV
MAAELFRAASMATOLOQETITPKPGTSLRQPVGVILSIVPMWNGPVLVLAARI
AYPLVCNVTIFRASMPRVHVAHVAAGFPQDSKLIVNLPEDAPALIEAMIAF
DRPVRVNTGSTRVGRIGELCGRHLKXSLVGLDGKSPVLVLCADYEAARVNAIFG
AFYQOQICMSTERLIVHETIADEFVERLAARVMTLKYGDPRQAQAAWALGPVVSILGA
DLNRLRLRLDLDDAVSGAKIAGGAAQQTENDATVLDHVTDRDAIYREEVFGPII"
32.0%; Score 105.6; DB 1; Length 6407;
Query Match
Best Local Similarity 58.7%; Pred. No. 2.6e-11;
```

ORIGIN

```
Matches 183; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
Qy 19 GTGATCTTCGCGCCAGCGCGGCTTCGAGCATCTGCTGAAACCGGAGGAGTATCG 78
Db 429 GTAAACATATTGAAACATGACGGAATTCAGCGCACCGTGGAGCGCATCCGGGTGCTCC 488
Qy 79 CTCATGGAAGCGCGCGTTCCTGAACCGCGTGGAGCGGTATCGAAGCCCGTTTCGCGGGCGCC 138
Db 489 ATCATGGAGATCGCGGTGAACACGCGGTTCGCGGTATTGATGCCGAATGTGGCGCGCC 548
Qy 139 TGTGCTTCGCGCCAGTGGCGACGTTTACGTTTGGCCCGCGAGTGGCTAGATGGCTGAAACCG 198
Db 549 TGGCGCTTCGCGCGACCTGCCACATCTATGTCGACGCGCGTGGCTTAAAGAACTTCCCGCG 608
Qy 199 CCGAGTGTAGACCGAAGACGAATGCTTCGATTGGTATGCGGAGCGAAGCGTGGCGCGCATTCGCGG 258
Db 609 CCGTCCGATTCCGAAGAGCGATGCTGGGATTCGCTGCGCATCGCCAGACGACACGCGC 668
Qy 259 CTGCTCTGCCAGATTCGCGCTTACCGACCTGCTGACGCGGCTGACCTGGAACCTGCCGAAG 318
Db 669 CTGCTCTGCCAGATTCGCGCTTACCGACCGAAGACGATGACGGAATCGTCTCAACGCGCGAA 728
Qy 319 GCACAGTCATGA 330
Db 729 TTCAGTTCTGA 740

RESULT 11
BX572605/c
LOCUS BX572605 349746 bp DNA linear BCT 18-DEC-2003
DEFINITION Rhodopseudomonas palustris CGA009 complete genome; segment 13/16.
ACCESSION BX572605 BX571963
VERSION BX572605.1 GI:39650627
KEYWORDS complete genomes.
ORGANISM Rhodopseudomonas palustris CGA009
Rhodopseudomonas palustris CGA009
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiacae; Rhodopseudomonas.
1 (bases 1 to 349746)
Larimer,F.W., Chain,P., Hauser,L., Lamerdin,J., Malfatti,S., Do,L.,
Land,M.L., Peltier,D.A., Beatty,T.J., Lang,A.S., Tabita,F.R.,
Gibson,J.L., Hanson,T.E., Torres y Torres,J., Pares,C.,
Harrison,F.H., Gibson,J. and Harwood,C.S.
Complete genome sequence of the metabolically versatile
photosynthetic bacterium Rhodopseudomonas palustris
Unpublished
2 (bases 1 to 349746)
Larimer,F.W. and Harwood,C.S.
Rhodopseudomonas genome consortium
Direct Submission
Submitted (24-JUL-2003) Submitted on behalf of the Rhodopseudomonas
genome consortium, the DOE Joint Genome Institute, Production
Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
larimerf@ornl.gov
Location/Qualifiers
1..349746
/organism="Rhodopseudomonas palustris CGA009"
/mol_type="genomic DNA"
/strain="CGA009"
/db_xref="taxon:258594"
96..521
/locus_tag="RPA3709"
96..521
/functions="InterPro IPR000971 COGS COG1017"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="possible hemoprotein"
/protein_id="CAE29150.1"
/db_xref="GI:39650628"
```

/translation="MMTPQAIALVQOSFARLTPISTDETATLTYERLFPATAPELRPLFH
 GDMRQCKLMLSTLVVVTGLTDLTTLTPATGLRLARLHVAGVSAHYTPFGAALLWT
 LERLGLAETWPELAAARDAYAMLSETMLAAAYGDTTPS"
 564..2336
 /gene="nirA"
 /locus_tag="RPA3710"
 564..2336
 /gene="nirA"
 /locus_tag="RPA3710"
 /EC_number="1.7.7.1"
 /function="InterPro
 IPR001424:IPR005117:IPR006066:IPR006067 COGs COG0155"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="possible ferredoxin-nitrite reductase"
 /protein_id="CAE29151.1"
 /db_xref="GI:39650629"
 /translation="MGSEFSIDQKRYLEGFATGVTAARAAGSLPAGSAASPGDPAI
 HQQDRVTVAGSKGLNQEKWRQHPFDAYLRKQAKDQNAAPKPADNFRWRYGLF
 YVAPQTSYMLCLIPNGILTSWQAGLAYLADTLADGYAHVTTTRANLQWREIAPDA
 VQLQGISLESGLWARGSGADNIRNVGTSGATGIDPQELIDTRAYARDWHEYILNERAL
 FGLPRKNVAFDGGGLIPTLEDNTDIAFOAVTIGDGHGAPGVVFRALGGLTGHKDF
 ARDGVVAPDEADAVIVRVEIEHGDRTDENKAEKLVLDARFGFDKYLEVEDKL
 GRKLIRVPAEAAAPRQDRAHLGVHPQQQAGNWNIGVRLATGHLTSTVUMRGLAIEIA
 RSPGQDRLTVWQNLISGVFDARVTDATAIEALGLAITASPLRAGLITACTGASGC
 RFAAHTKETAEIAQYCEBPRLDTPINHLTGCHNSCAQHFISDGLIGAKVAVSD
 EDSVEGHHVGGFGVDAAIGTEVLRDVKQDDAPRVIERMLOTYLARRASPAETFLA
 FVRQDAALQQLPAYEQSPETVA"
 2333..3952
 /locus_tag="RPA3711"
 2333..3952
 /locus_tag="RPA3711"
 /EC_number="1.8.1.2"
 /function="InterPro IPR001433:IPR001709:IPR003097 COGs
 COG0369"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="possible sulfite reductase (NADPH)"
 /protein_id="CAE29152.1"
 /db_xref="GI:39650630"
 /translation="MSNLOPPIPLLVPEYAPFSEQRWLNGFFAGVLSDQGGVTA
 LSQQAALLAGGAPAEQEDDGGAPWHQDSLPIGRMQLABGKPLRWLMAAQQD
 CGQVDCRNYVSGALFQGOETRNLNLCAPGKDTARMVKTLTTEOIGRAPTLQDAPSLAA
 DAAPVVALPRTGTRDNPATKVLRSKLNKPGSEKETWVUELEDCLSYEAGDSFG
 LFTNDPALVDVHALGAPAEFFIAQSTLRQLTLDVSVLSLPADMLFQISITIGD
 RKKARSASGEDPDGAATLDVLALEKFPGLRPDPPEAFIADLPQPLRFLYISDPN
 KTTPCRSLTVDCVRYAIGKQRLGVCSTGLAERVREGTIRAYVQKARHFLYISDPN
 QPIIMIGPTGIAPFLHERQAIQAPGNLWFFGHQSGASDFYDEDELKAMKTAGH
 LTRLTAWSRDSGEKIVQDRIREVGRDLMSWLAEGASLYVCGDAKEMAKDVERALVD
 IVAOHGARSAAEATAFVSELKQGRYQQDVT"
 complement(4092..4863)
 /gene="pimE pseudogene"
 /locus_tag="RPA3712"
 /pseudo
 complement(join(4092..4487,4486..4863))
 /gene="pimE pseudogene"
 /locus_tag="RPA3712"
 /EC_number="1.1.1.1"
 /function="InterPro IPR002198 COGs COG1028"
 /note="frameshift"
 /pseudo
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="short chain dehydrogenase"
 /db_xref="PSEUDO:CAE29153.1"
 complement(5013..6155)
 /gene="pimD"
 /locus_tag="RPA3713"
 complement(5013..6155)
 /gene="pimD"

/locus_tag="RPA3713"
 /EC_number="1.3.1.62"
 /function="InterPro IPR001687:IPR006090:IPR006091 COGs
 COG1960"
 /note="observed by proteomics
 Citation: Proteomics from VerBerkmoes et al. (2003)
 unpublished"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="pimeloyl-CoA dehydrogenase (small subunit)"
 /protein_id="CAE29154.1"
 /db_xref="GI:39650631"
 /translation="MDFDLSBEQRLLKESVEGLLKGSYDFDSRKYAKKSGWSRAVW
 GKFAEQGLLGLPFSBEDGGFGAGAVETMVMVMEALGHSVLBPYLPVTVVGGFLLRAG
 SAAOKAAHLPGIIGDSKTFAPAOLEKNSRWLDGVSTAKSGDGDWIDGKFKVVLNG
 EAAADTLVYARTKGSQDRDTGVGVFLVPADAKGITRKGYPTQDGLHAADIITFGVQVG
 ADAAIGDPENALLEIEAVDDARTALCAEAVGLMDESITTVYEIKTRKQGVPIGS?
 QVLOHRAADMVATEQASMAFMATMAAEFDDAKERAGAAIAAKVQIGSKGRFVCGQS
 IQLHGIGMTWEAKIGHYFKLTMTEQTFGDTDHLARVSAGGGLI"
 complement(6324..7514)
 /gene="pimC"
 /locus_tag="RPA3714"
 complement(6324..7514)
 /gene="pimC"
 /locus_tag="RPA3714"
 /EC_number="1.3.1.62"
 /function="InterPro IPR006090:IPR006091:IPR006092 COGs
 COG1960"
 /note="observed by proteomics
 Citation: Proteomics from VerBerkmoes et al. (2003)
 unpublished"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="pimeloyl-CoA dehydrogenase (large subunit)"
 /protein_id="CAE29155.1"
 /db_xref="GI:39650632"
 /translation="MDLNFPSKEEIAFRDEVQFQFKNVPAKTRKLIQRHNTKEEMV
 EWTIRLNKGNWAVTHWPXEYQSGVSSVCHYIFNEELOAAPQAPLAFGVSNGVPIY
 TFGSEBQKRLPLRIANVDDWMCQFSGFSGSDLSLKTAEKRGKDWINGQKTWT
 TLAADHIVFCLRTDPAKQEGISFILVPMKTKGITVTRPIQTTIDGHEVNEVFDD
 VEVPLENVGKNGWYAKFLQGNERTGIARVGNKERRIRIKQLAAQVSEGGKPI
 EDPKRGDLAAVEIELKALELTQLRVADGEGKGNPKNPASSVLKIKGSEIQOATTE
 LMVEYFGPFAAPYDVHGDDSDNTWDTAQIAPGYFNNKRVSIYCGSIQNRNICKA
 VLGL"
 complement(7729..8916)
 /gene="pimB"
 /locus_tag="RPA3715"
 complement(7729..8916)
 /gene="pimB"
 /locus_tag="RPA3715"
 /EC_number="2.3.1.9"
 /function="InterPro IPR002155 COGs COG0183"
 /note="observed by proteomics
 Citation: Proteomics from VerBerkmoes et al. (2003)
 unpublished"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="acetyl-CoA acetyltransferase"
 /protein_id="CAE29156.1"
 /db_xref="GI:39650633"
 /translation="MTEAVIVSTARTPIGKAYRGALNATEGATLGLHAIEHAVKRAGI
 SVLPFDGVLAIVGGGESISLVNDKMTFHAVDPALEAIKGVVYVMDLTDAETVAKRY
 DPKEVEDVMGAAMQCGATGGNIARAKALLRAGLPVTAGTITDRCQASGLQALAAAR
 SVLPFDGVLAIVGGGESISLVNDKMTFHAVDPALEAIKGVVYVMDLTDAETVAKRY
 GTSRRQDEYLSLESORRTAAQGGKFNDETAPISTKMGVVDKATGAVSPKIDILSOD
 EGPRTTAEGLAGLKAIVRGEGFTTAGNSQLSDGASATVIMSRTAAKGLKPLGI
 FRGMVSYGCEPDGEMIGIPFAVFAVPRLLKRLHSLGVSDDIGLWELNEAFVQVLYCRDKLGI
 DPEKLVNNGAISVGHPIYGMGARGALHGLERRRKRAKIAVVTVMVCGGMSAGLFE
 IVH"
 complement(8948..10606)
 /gene="pimA"


```

CDS
complement(3955..5025)
/note="CE0293, similar to AX065881-1|CAC26180.1| percent
identity: 56 in 305 aa"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAC17103.1"
/db_xref="GI:23492129"
/translation="MTTRTSPGSPVHADQSPSPSPSRMVKRFTPEPWKLWMTA
LRRRSVNRASERTPTVAVEGYPYRHDVYVGRVLRHVAVAGSPADPLVLLHGA
FGFVDFYDVTGSLAAWFFVAAVDLGYGLSDPKPSGYDIRRSAGDINGVIGLALGH
DAIIVGDTGSLAAWFFVAAVDLGYGLSDPKPSGYDIRRSAGDINGVIGLALGH
LFSPIPLVASIRPVPRASREVTRCTTASVORSNAYTCATLRKALALDHTFPEI
VTRNELTASLPGSDHOLARCPVLLTPRDNTEYLASLARSRTSGPFLVSIPIGT
ALSTLEDPLETFEMIASFLHDS"
complement(5239..5895)
/note="CE0294, similar to AB001940-6|AAF10446.1| percent
identity: 23 in 150 aa"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAC17104.1"
/db_xref="GI:23492130"
/translation="MARLGLKNAQVIRLKEHWSFQVNNNDGLFTDGTETFSKVNAI
PLSDVTSISGEASIGTILSNATQMSLSFRAEVELAKELAGEAKKAAGGAFGAA
GVIALYSPFPFLASTLKIWLDMSLYLIVLEMLIAAVALPGWKVKWKGAPK
KTIDSVNELNLVPGQATSNLERSDKRLYTSNFPAGSTAASTIPSGHGTAYRK"
complement(6235..7683)
/note="CE0295"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC17105.1"
/db_xref="GI:23492131"
/translation="MSGGSDGLPPELTQRPVAPDPVAPDKRPAFNPVHVMLLP
PSLDEWFLDNHARFIADVLDELDTGFDYSYTRSKRPPYDFRMLRVMLYGYCTG
VRSRKLFTICISVAFRLAAGDPFRFARHRLQALAGVFVQALALCREAGM
VLTGVALDGTQVANSRRKMSYQRLVPAEKLAEQVEMLNDAATDEVDQRYG
VDARGDELPELVRSRLATLRARQOLEODAAKARQVATKARDRGDDDTVDVK
QAAADTAVPMVLATAEOLVMPAELMDAGYCKNLTVEAGLEDHTFEFFATGSEV
GDAAGTAVPMVLATAEOLVMPAELMDAGYCKNLTVEAGLEDHTFEFFATGSEV
KHGSKIIPDAFRGRIIPNATLRERMGRRLRYKRGKAVYKRSVIEPVFGQIATRQGH
VVLKGLAAARAEVDLVAGCHNLLKLSFRQTA"
complement(7804..8850)
/note="CE0296, similar to AJ010584-1|CAB55344.1| percent
identity: 31 in 273 aa"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAC17106.1"
/db_xref="GI:23492132"
/translation="MLGEWVDRGAERGWLHVI VNSPPQGGRIEPTSTRNHTSHR
NAGLPLIHRKTRVTCVTPNDFSPGHFNAAQDSTRAAPRRVAFVFLDXTIIM
SSTVAYGREFMONGLISPVTAQLSLAQATVMPAGHSTSEOMDTRDOLTAMVGNVQ
OVRTAETWCTVGTPTIYAEARELDYHREGRHDVILIASVKEAVEPTARELVNQ
TVSTLTIDGRYGEVLFYCKGAQQAQVNLADQGYDESSYAYSADATDIPMD
IGVHPVAVNPDRALKKAALERGDWILSKFNPFLQFMPSTREMGIGTVGVVGAIAVTA
GGHWARRARRGSA"
complement(8903..10303)
/note="CE0297"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC17107.1"
/db_xref="GI:23492133"
/translation="MQGPDILCGLFRPPEVGLPTARHPPRGAGRPHPSPSSIGPL
PGGFLPGRRPGPGYPPGSPSTKPGVTHRLAASLRAGPAASRWGMGNATTTTTR
TAGTARSTQDOAILVAVBDPVLVHPAMHVAATGTAIVDTTDPVDIRHLLKYSA
VLIDAPTAGITNDRERDRIELGSGPQPDYHTALAIQAQALLPAQTAELLQVLA
REDSAPGRHATVTGVLGAGGVGVSTIAALARISATHTVLVDVPTSGGIDL
LVGABETGARWPDGLTRGAVKAEVDVLAALPVMDETMFIUSGARSVPVGTDFDLPDE

LQCAVS"
complement(3955..5025)
/note="CE0293, similar to AX065881-1|CAC26180.1| percent
identity: 56 in 305 aa"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAC17103.1"
/db_xref="GI:23492129"
/translation="MTTRTSPGSPVHADQSPSPSPSRMVKRFTPEPWKLWMTA
LRRRSVNRASERTPTVAVEGYPYRHDVYVGRVLRHVAVAGSPADPLVLLHGA
FGFVDFYDVTGSLAAWFFVAAVDLGYGLSDPKPSGYDIRRSAGDINGVIGLALGH
DAIIVGDTGSLAAWFFVAAVDLGYGLSDPKPSGYDIRRSAGDINGVIGLALGH
LFSPIPLVASIRPVPRASREVTRCTTASVORSNAYTCATLRKALALDHTFPEI
VTRNELTASLPGSDHOLARCPVLLTPRDNTEYLASLARSRTSGPFLVSIPIGT
ALSTLEDPLETFEMIASFLHDS"
complement(5239..5895)
/note="CE0294, similar to AB001940-6|AAF10446.1| percent
identity: 23 in 150 aa"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAC17104.1"
/db_xref="GI:23492130"
/translation="MARLGLKNAQVIRLKEHWSFQVNNNDGLFTDGTETFSKVNAI
PLSDVTSISGEASIGTILSNATQMSLSFRAEVELAKELAGEAKKAAGGAFGAA
GVIALYSPFPFLASTLKIWLDMSLYLIVLEMLIAAVALPGWKVKWKGAPK
KTIDSVNELNLVPGQATSNLERSDKRLYTSNFPAGSTAASTIPSGHGTAYRK"
complement(6235..7683)
/note="CE0295"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC17105.1"
/db_xref="GI:23492131"
/translation="MSGGSDGLPPELTQRPVAPDPVAPDKRPAFNPVHVMLLP
PSLDEWFLDNHARFIADVLDELDTGFDYSYTRSKRPPYDFRMLRVMLYGYCTG
VRSRKLFTICISVAFRLAAGDPFRFARHRLQALAGVFVQALALCREAGM
VLTGVALDGTQVANSRRKMSYQRLVPAEKLAEQVEMLNDAATDEVDQRYG
VDARGDELPELVRSRLATLRARQOLEODAAKARQVATKARDRGDDDTVDVK
QAAADTAVPMVLATAEOLVMPAELMDAGYCKNLTVEAGLEDHTFEFFATGSEV
GDAAGTAVPMVLATAEOLVMPAELMDAGYCKNLTVEAGLEDHTFEFFATGSEV
KHGSKIIPDAFRGRIIPNATLRERMGRRLRYKRGKAVYKRSVIEPVFGQIATRQGH
VVLKGLAAARAEVDLVAGCHNLLKLSFRQTA"
complement(7804..8850)
/note="CE0296, similar to AJ010584-1|CAB55344.1| percent
identity: 31 in 273 aa"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAC17106.1"
/db_xref="GI:23492132"
/translation="MLGEWVDRGAERGWLHVI VNSPPQGGRIEPTSTRNHTSHR
NAGLPLIHRKTRVTCVTPNDFSPGHFNAAQDSTRAAPRRVAFVFLDXTIIM
SSTVAYGREFMONGLISPVTAQLSLAQATVMPAGHSTSEOMDTRDOLTAMVGNVQ
OVRTAETWCTVGTPTIYAEARELDYHREGRHDVILIASVKEAVEPTARELVNQ
TVSTLTIDGRYGEVLFYCKGAQQAQVNLADQGYDESSYAYSADATDIPMD
IGVHPVAVNPDRALKKAALERGDWILSKFNPFLQFMPSTREMGIGTVGVVGAIAVTA
GGHWARRARRGSA"
complement(8903..10303)
/note="CE0297"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC17107.1"
/db_xref="GI:23492133"
/translation="MQGPDILCGLFRPPEVGLPTARHPPRGAGRPHPSPSSIGPL
PGGFLPGRRPGPGYPPGSPSTKPGVTHRLAASLRAGPAASRWGMGNATTTTTR
TAGTARSTQDOAILVAVBDPVLVHPAMHVAATGTAIVDTTDPVDIRHLLKYSA
VLIDAPTAGITNDRERDRIELGSGPQPDYHTALAIQAQALLPAQTAELLQVLA
REDSAPGRHATVTGVLGAGGVGVSTIAALARISATHTVLVDVPTSGGIDL
LVGABETGARWPDGLTRGAVKAEVDVLAALPVMDETMFIUSGARSVPVGTDFDLPDE

VTALTCTAAGDELEVVDLNPGEITREVIPLDLHLILVVPVPAEVAVAAAEERHUL
RAFPVPSVLRHRCWGLSDIEVERILGTVEIAELGTITRPLRAVEMHGLTGLPRP
LVTGNAIAAEIRGRG"
complement(10303..12111)
/note="CE0298, similar to AL035636-18|CAB38493.1| percent
identity: 55 in 348 aa"
/codon_start=1
/transl_table=1
/product="putative conjugal transfer protein"
/protein_id="BAC17108.1"
/db_xref="GI:23492134"
/translation="MSTTIPDELEQVORVLAGLPEHPTSPVVOVIREQAGVISDE
EVALRLRRESVGIGLEAALAPGVTVLVNGPRVWIDRGAGLSQLDLGSESE
AVRLAARLALQCPRLDAPQFVDGCTTRDGSITRHAIVPLADSGTCLSMELLR
QACITLDLIVAGGTLTGTRAVLRITVAQRRAFLVVGTSKTKLLAMLEGVAPDQ
RICTEDTAELNPPHPTGNLVITRSANVEGAGATMSDLLKQAMRPRDRIVLGEIR
AEVVELLAALNTHEGGAGTTHANSIAEVPARMEALALGLADLHLSQALAAADV
IVMHTTTRRLAQLGLVLRIPSRMWCCTGTAWMTQWRHGFRTDPPFGRDGA
CAGAGGTGRRASGSGTPATCGHRRRPRVRGPRVRRHRRPRHGGGAADRRTRGELA

Query Match 27.0%; Score 89; DB 1; Length 309350;
Best Local Similarity 60.4%; Pred. No. 3.5e-08;
Matches 165; Conservative 0; Mismatches 105; Indels 3; Gaps 1;

QY 25 TTCGCGCAGCGCGCGCTTCGAGCATCTGGTCGAAACCGAGCGGAGTAGTCGCTCATG 84
Db 278820 TTCACCGATACCACCGCGAGACCCGACCATCAACGCAACGCTCCGCTGATG 278761
QY 85 GAAGCGCGCTTCTGACCGCGTGCAGGATATCGAAGCGCTTTCGGGGCGCGCTGTGCC 144
Db 278760 GAGACCGCGCTGCGCAACGGTGTGCGGGCATCGTCGCGCAATGCGGGCGTTCCTGTCC 278701
QY 145 TCGCCACGTCGCGCATTTTACCTTGGCCCGAGTGGCTAGATCGCTGAAACCGCGCGAGT 204
Db 278700 TCGCCACGTCGCGCATTTTACCTTGGCCCGAGTGGCTAGATCGCTGAAACCGCGCGAGT 278641
QY 205 GAGACGGAAGCAAAATGCTCGATTGCGTAGC---GGAACGTGCGCGCATTCGCGGCTG 261
Db 278640 GAGATGAGAGCAGAGATGCTCTGGGGCGCGCGCTGAGACCGAGAGTGTTCGCGCTG 278581
QY 262 TCTGCGCAGATCGCGCTTACCGACCTGCTCGAC 294
Db 278580 TCTGCGCAGATCGCGTTCACCGAGCGCATG 278548

RESULT 13
LOCUS
DEFINITION
AB008062 10495 bp DNA linear BCT 18-DEC-2001
Agrobacterium tumefaciens str. C58 circular chromosome, section 120
of 254 of the complete sequence.
ACCESSION
AB008062 AB007869
VERSION
AB008062.1 GI:15156410
KEYWORDS
ORGANISM
Agrobacterium tumefaciens str. C58 (Cereon)
Agrobacterium tumefaciens str. C58 (Cereon)
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
1 (bases 1 to 10495)
REFERENCE
Hinkle,G., Slater,S.C. and Goodner,B.
Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
Unpublished
JOURNAL
2 (bases 1 to 10495)
AUTHORS
Hinkle,G., Slater,S.C. and Goodner,B.
TITLE
Direct Submission
JOURNAL
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA
FEATURES
Location/Qualifiers
1..10495
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/mol_type="genomic DNA"
/strain="C58"

```



```

/product="AGR_C 2511p"
/protein_id="AAK87150.1"
/db_xref="GI:15156420"
/translation="MTASFTRVYLLRHAAWAAGGERDFRGLNEAGFAEELIAD
LAADRRPDLILSSTAARCTQAWRAFNGEIDIVIDENYNAKSETIYLSLIAAQ
TEQSVMLVGHNTTMEATLEAMIGEDLLHAALFSGFPTSLGLVLDQDDSAASGNRWR
LIDFLAFPK"
8082..8528
/gene="AGR_C 2513"
8082..8528
/gene="AGR_C 2513"
/notes="DNAK SUPPRESSOR PROTEIN HOMOLOG"
/codon_start=1
/transl_table=11
/product="AGR_C 2513p"
/protein_id="AAK87151.1"
/db_xref="GI:15156421"
/translation="MRCSSRRIALSEKIDSKVLSDEDFMASORAYPRAKLIAWK
NDILREARELIGHLAESAHPDLADRASSETDRIELRARDQRKLKIDLAALQRI
DDGTGYCEETGFEIGLGRUDAPRIATLSIEAQRHERREKVRDE"
complement(8637..9887)
/gene="AGR_C 2517"
complement(8637..9887)
/gene="AGR_C 2517"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C 2517p"

Query Match      26.1%; Score 86.2; DB 1; Length 10495;
Best Local Similarity 59.0%; Pred. No. 1.8e-07;
Matches 148; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 80 TCATGGAGCGCGTTCGTGACCGCGTGGACGATATCGAACCGTTTGGCGGGCGCCT 139
DB 829 TCATGGAGATGCGGTGCGCAACTCCGTTCGCCGATGTATGCGGAATGCGGCGCGCCT 888
QY 140 GTGCTGCGCCAGCGGCACGCTTACGTGTGCGCCGAGTGTAGATCGGTGAACCGC 199
DB 889 CGCCTGCGCCACCTGCCAGCTGTATGTTCAGATGCTGGCAGAGCGTGTGGCGGAC 948
QY 200 CGAGTGACCGAGACGAATGCTCGATTCGATGCGGAGAACGTTGCGCGCGATTCGCGGC 259
DB 949 CCGAGCCGATGGAAGAGACATGCTGGCACTTCGCGCTTCGAGGTTCTGCCCACTTCGCGGC 1008
QY 260 TGCTCTGCAGATCGCGCTTACGACCGTCTGACGCGCTCACCTCGAACTGCGGAAGG 319
DB 1009 TTTCCTGTCAGATCAGATGATGACGAGCTTGACGGTCTCGTTGTCATGTTCCCGAAC 1068
QY 320 CACAGTCATGA 330
DB 1069 GCCAGGCTGA 1079

```

```

RESULT 14
AE009097
LOCUS
DEFINITION
  Agrobacterium tumefaciens strain C58 circular chromosome, section
  123 of 256 of the complete sequence.
ACCESSION
  AE009097 AE008688
VERSION
  AE009097.1 GI:17739761
KEYWORDS
  Agrobacterium tumefaciens str. C58 (U. Washington)
  Agrobacterium tumefaciens str. C58 (U. Washington)
  Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
  Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
  1 (bases 1 to 10683)
REFERENCE
  AUTHORS
    Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
    Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
    Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,
    Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,
    McClelland, E., Palmieri, A., Raymond, C., Rouse, G.,
    Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
    Krespan, W., Perry, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
    and Nester, E.W.
  The genome of the natural genetic engineer Agrobacterium
  tumefaciens C58
  Science 294 (5550), 2317-2323 (2001)

```

```

TITLE
  The genome of the natural genetic engineer Agrobacterium
  tumefaciens C58
  Science 294 (5550), 2317-2323 (2001)
JOURNAL
  MEDLINE
  PUBMED
REFERENCE
  AUTHORS
    Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,
    Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
    Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
    and Nester, E.W.
  The genome of the natural genetic engineer Agrobacterium
  tumefaciens C58
  Science 294 (5550), 2317-2323 (2001)
  2 (bases 1 to 10683)
    Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
    Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
    Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,
    Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,
    McClelland, E., Palmieri, A., Raymond, C., Rouse, G.,
    Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
    Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,
    Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
    Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
    and Nester, E.W.
  Direct Submission
  Submitted (27-SEP-2001) Department of Microbiology, University of
  Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
  98195-7242, USA
FEATURES
  source
    1..10683
    /organism="Agrobacterium tumefaciens str. C58 (U.
    Washington)"
    /mol_type="Genomic DNA"
    /strain="C58"
    /db_xref="taxon:180835"
    209..577
    /gene="Atu1349"
    209..577
    /gene="Atu1349"
    /note="putative; ORF located using Glimmer"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA142355.1"
    /db_xref="GI:17739762"
    /translation="MAAVSIVFEAPDNYGPAVSKPTDPEHLAQOTMGDKLEIEV
    LQFRRSAPALHHEAGADNKVTATRAHLKGSAGVAGASAAAAYEKGNDAL
    IAKLAAIVEAENFILKLR"
    707..1027
    /gene="fdx"
    /note="synonym: Atu1350"
    707..1027
    /gene="fdx"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer"
    /codon_start=1
    /transl_table=11
    /product="ferrodoxin"
    /protein_id="AA142356.1"
    /db_xref="GI:17739763"
    /translation="MTKLTIAPDGTPEHLDVSGSTWENAVRNVSPGIDACCGAC
    ACATCHVYDDAWAEVRVGPPEEDMLDFAFEVPTSLSCQIKNDELGLGVHVP
    ERQG"
    complement(1334..1960)
    /gene="Atu1351"
    complement(1334..1960)
    /gene="Atu1351"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AA142357.1"
    /db_xref="GI:17739764"
    /translation="MTKSSSALCAVFAAGALSLSGPAFAFTIVKTVSYFVSQKTA
    ADDQQLSKHGLTTRNTGARHFGATEIKFGGLITVYKNGMSGVTAFLVLRILLP
    QWKRRTTAEALFTWDTLLADIKRHEERHAEIARTHAKSLERQUTSLRFPQKSLVQ

```


	Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
REFERENCE	1 (bases 1 to 306250)
AUTHORS	Capeila,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J., Boitard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S., Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D., Pohl,T., Portetlelle,D., Puchelot,A., Purnelle,B., Ramsperger,U., Renard,C., Thebaud,P., Vandenberg,M., Weidner,S. and Galibert,F. Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021 Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001) 21396507 11481430
TITLE	Gouzy,J.
JOURNAL	MELILO EU Consortium:
MEDLINE	Laboratoire de Biologie Molculaire des Relations
PUBMED	Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet, France, Laboratoire de Genetique et Developpement UMR6061-CNRS, Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes, France, GARC GmbH, Fritz-Arnold-Str. 23, D-78467 Konstanz, Germany, Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany, Unite de Biochimie physiologique, Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20, B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6, B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr http://sequence.toulouse.inra.fr/meliloti.html . Location/Qualifiers
FEATURES	1..306250 /organism="Sinorhizobium meliloti" /mol_type="genomic DNA" /strain="1021" /db_xref="taxon:382"
source	488..2707 /gene="SMC00954" 488..2707 /gene="SMC00954" /function="miscellaneous; hypothetical/global homology"
Gene	/note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Homology predicted by Framed"
CDS	/codon_start=1 /evidence=not experimental /transl_table=11 /product="CONSERVED HYPOTHETICAL PROTEIN" /protein_id="CAC46225.1" /db_xref="GI:15074580" /db_xref="GOA:Q92PT4" /db_xref="SPTRMBL:Q92PT4"
	/translation="MPAQPHLPANAPNIVIMLDVGPALPDTFGGVIETPTLSRLAQ DGLSYNRFNAMGSPSAALLTGRNHHVGNGQIAELANDMGYTCRTPRTSATYAK VLGYGYATAPAGKWNHPANNETTAVGTWDPAIGEGIDFYFGLAGESSQWEPAY VENTVRVPSHGKEGTHETDMTNKAVKQMVKQHALTPDRFFWMYAPGAAPHPIIF KDWAQKQDDGDGDAWRERIYARQKLGWI PQNTQLTPRPDTLSGMDIPDEKPF ORLMVEFAGTYHADVAQKLLEALDELGI RDNTLI FYVMGNGSABEQONTISEL LAQNGIRTEIKDHIRANMELGMDVLPKADNMHAGWAGSTPHRSKTLAAHFGL GRTPLPKVPCKIKPDOTPRSOFHLNDIVPTIYVDLDIQPKLVNDGISQDLGVIS MTYTDSPEASQKEQOEYFMGSRAIYQEWIASVFGRPKALDPAIFKWSDN DLWEHLSDQSQXONKNAADHPKFKALGFAGVQAENKVPFGGGLMSAFPHED APSNPATBFQTQDVGVGEFTAPKVARSNLITAELEPDSSEGLVALGAFSGGVA LWVENGLITYNFYEIRTERLESDDLPGSKVNIYEYTRKVSNSDHAAPLDVVIRIDG KEVAKGHVPRSPAFLTAFTANDAFDGKDSYSVALAYFDKPPAFNGKIBSLKVEYLK" 2887..4062 /gene="SMC00955" 2887..4062 /gene="SMC00955" /function="miscellaneous; hypothetical/global homology"
Gene	/note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Homology predicted by Framed"
CDS	/codon_start=1 /evidence=not experimental /transl_table=11 /product="CONSERVED HYPOTHETICAL PROTEIN" /protein_id="CAC46225.1" /db_xref="GI:15074580" /db_xref="GOA:Q92PT4" /db_xref="SPTRMBL:Q92PT4"
	/translation="MKIATYNVNGVRGLRVLRUEBASPDVCLQEUKADDPKFPY KATEAAGYAIGHQKSWNGVAIALAREFTTKRGLPDGDPTHSHRYTEAAVEGVKI GCILYPNGNYPGPKFEYKLAWFRLHTAYAAELLELDVPLIAGDNNVMMPTSLDYVK ERWYDALPIEVRDAYRLLIEQGWTDLRQLHPGSRVTTFWDYFRNAPFADAGLRID HLILSPHVTRLISAAGVDHRVRCWENTSHPAPIELSOGPAEDQ" 5519..5737 /gene="SMC00958" 5519..5737 /gene="SMC00958" /function="miscellaneous; unknown"
Gene	/note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Homology predicted by Framed"
CDS	/codon_start=1 /evidence=not experimental /transl_table=11 /product="PROBABLE EXODEOXYRIBONUCLEASE III PROTEIN" /protein_id="CAC46228.1" /db_xref="GI:15074583" /db_xref="GOA:Q92PT1" /db_xref="SPTRMBL:Q92PT1"
	/translation="MKIATYNVNGVRGLRVLRUEBASPDVCLQEUKADDPKFPY KATEAAGYAIGHQKSWNGVAIALAREFTTKRGLPDGDPTHSHRYTEAAVEGVKI GCILYPNGNYPGPKFEYKLAWFRLHTAYAAELLELDVPLIAGDNNVMMPTSLDYVK ERWYDALPIEVRDAYRLLIEQGWTDLRQLHPGSRVTTFWDYFRNAPFADAGLRID HLILSPHVTRLISAAGVDHRVRCWENTSHPAPIELSOGPAEDQ" 5519..5737 /gene="SMC00958" 5519..5737 /gene="SMC00958" /function="miscellaneous; unknown"
Gene	/note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Homology predicted by Framed"
CDS	/codon_start=1 /evidence=not experimental /transl_table=11 /product="HYPOTHEITICAL/UNKNOWN PROTEIN" /protein_id="CAC46227.1" /db_xref="GI:15074582" /db_xref="SPTRMBL:Q92PT2"
	/translation="MSKRELIDGTDKRYVRDESCKFKHSVDVGRSLSSDKKQAKAK DAPKGQDKGRKN" 4728..5522 /gene="xthal1 OR SMC00956" 4728..5522 /gene="xthal1 OR SMC00956" /EC_number="3.1.11.2"
Gene	/function="macromolecule metabolism; macromolecule degradation; degradation of dna" /note="Product confidence : probable Gene name confidence : probable predicted by Homology predicted by Framed"
CDS	/codon_start=1 /evidence=not experimental /transl_table=11 /product="HYPOTHEITICAL/UNKNOWN PROTEIN" /protein_id="CAC46227.1" /db_xref="GI:15074582" /db_xref="SPTRMBL:Q92PT2"
	/translation="MSKRELIDGTDKRYVRDESCKFKHSVDVGRSLSSDKKQAKAK DAPKGQDKGRKN" 4728..5522 /gene="xthal1 OR SMC00956" 4728..5522 /gene="xthal1 OR SMC00956" /EC_number="3.1.11.2"
Gene	/function="macromolecule metabolism; macromolecule degradation; degradation of dna" /note="Product confidence : probable Gene name confidence : probable predicted by Homology predicted by Framed"
CDS	/codon_start=1 /evidence=not experimental /transl_table=11 /product="HYPOTHEITICAL/UNKNOWN PROTEIN" /protein_id="CAC46227.1" /db_xref="GI:15074582" /db_xref="SPTRMBL:Q92PT2"
	/translation="MSKRELIDGTDKRYVRDESCKFKHSVDVGRSLSSDKKQAKAK DAPKGQDKGRKN" 4728..5522 /gene="xthal1 OR SMC00956" 4728..5522 /gene="xthal1 OR SMC00956" /EC_number="3.1.11.2"
Gene	/function="macromolecule metabolism; macromolecule degradation; degradation of dna" /note="Product confidence : probable Gene name confidence : probable predicted by Homology predicted by Framed"
CDS	/codon_start=1 /evidence=not experimental /transl_table=11 /product="HYPOTHEITICAL/UNKNOWN PROTEIN" /protein_id="CAC46227.1" /db_xref="GI:15074582" /db_xref="SPTRMBL:Q92PT2"
	/translation="MSKRELIDGTDKRYVRDESCKFKHSVDVGRSLSSDKKQAKAK DAPKGQDKGRKN" 4728..5522 /gene="xthal1 OR SMC00956" 4728..5522 /gene="xthal1 OR SMC00956" /EC_number="3.1.11.2"
Gene	/function="macromolecule metabolism; macromolecule degradation; degradation of dna" /note="Product confidence : probable Gene name confidence : probable predicted by Homology predicted by Framed"
CDS	/codon_start=1 /evidence=not experimental /transl_table=11 /product="HYPOTHEITICAL/UNKNOWN PROTEIN" /protein_id="CAC46227.1" /db_xref="GI:15074582" /db_xref="SPTRMBL:Q92PT2"
	/translation="MSKRELIDGTDKRYVRDESCKFKHSVDVGRSLSSDKKQAKAK DAPKGQDKGRKN" 4728..5522 /gene="xthal1 OR SMC00956" 4728..5522 /gene="xthal1 OR SMC00956" /EC_number="3.1.11.2"
Gene	/function="macromolecule metabolism; macromolecule degradation; degradation of dna" /note="Product confidence : probable Gene name confidence : probable predicted by Homology predicted by Framed"
CDS	/codon_start=1 /evidence=not experimental /transl_table=11 /product="HYPOTHEITICAL/UNKNOWN PROTEIN" /protein_id="CAC46227.1" /db_xref="GI:15074582" /db_xref="SPTRMBL:Q92PT2"
	/translation="MSKRELIDGTDKRYVRDESCKFKHSVDVGRSLSSDKKQAKAK DAPKGQDKGRKN" 4728..5522 /gene="xthal1 OR SMC00956" 4728..5522 /gene="xthal1 OR SMC00956" /EC_number="3.1.11.2"
Gene	/function="macromolecule metabolism; macromolecule degradation; degradation of dna" /note="Product confidence : probable Gene name confidence : probable predicted by Homology predicted by Framed"
CDS	/codon_start=1 /evidence=not experimental /transl_table=11 /product="HYPOTHEITICAL/UNKNOWN PROTEIN" /protein_id="CAC46227.1" /db_xref="GI:15074582" /db_xref="SPTRMBL:Q92PT2"
	/translation="MSKRELIDGTDKRYVRDESCKFKHSVDVGRSL

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2004, 03:08:33 ; Search time 144.642 Seconds
(without alignments)
9692.234 Million cell updates/sec

Title: US-10-658-691-2
Perfect score: 330
Sequence: 1 atgacgcgaagtcgcgct.....tgcgaagcacaagtcagta 330

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	74	22.4	8766	7 ACC48896	Acc48896 Rhodococc
2	74	22.4	12342	7 ACC48895	Acc48895 Rhodococc
3	74	22.4	23656	7 ACC48894	Acc48894 Rhodococc
C 4	65.6	19.9	306	5 AAH65575	AAH65575 C glutami
5	65.6	19.9	318	5 AAH65577	AAH65577 C glutami
6	65.6	19.9	441	4 AAF71697	AAf71697 Corynebac
C 7	65.6	19.9	34980	5 AAH68525	AAH68525 C glutami
8	62.8	19.0	389	4 ABL07331	ABl07331 Drosophil
9	62.8	19.0	2389	4 ABL07330	ABl07330 Drosophil
10	62.8	19.0	4531	4 ABL07682	ABl07682 Drosophil
11	62.8	19.0	4531	4 ABL07688	ABl07688 Drosophil
12	60.4	18.3	458	8 ACH38677	Ach38677 Human fo
13	60.4	18.3	473	8 ACH33408	Ach33408 Human end
14	60.4	18.3	834	6 ABQ60909	ABq60909 Signal pe
15	60.4	18.3	840	3 AAZ98240	Aaz98240 Human sig
16	60.4	18.3	958	4 AAD16583	Aad16583 Human nov
17	60.4	18.3	958	4 AAD16590	Aad16590 Human nov
18	60	18.2	632	7 ABZ36606	Abz36606 Human GEN
19	60	18.2	1166	7 ABZ36446	Abz36446 Human GEN
20	58.2	17.6	413	7 ABX42738	Abx42738 Bovine ES
21	58.2	17.6	443	7 ABX49100	Abx49100 Bovine ES
22	57.6	17.5	1950	9 ADE73150	Ade73150 Pseudomon
23	57.6	17.5	2232	2 AAQ13693	Aaq13693 Putidared

24	54.4	16.5	300	3 AAC30981	Aac30981 Human sec
25	44	13.3	612	6 ABZ32192	Abz32192 Candida a
26	43.8	13.3	393	7 ACD95816	Acd95816 Human col
27	43	13.0	508	5 AAS64988	Aas64988 DNA encod
28	42.2	12.8	44377	2 AAT80414	Aat80414 Platenoli
29	42.2	12.8	44377	2 AAT80414	Aat80414 Platenoli
30	41.8	12.7	1501	5 AAS91987	Aas91987 DNA encod
31	41.8	12.7	1506	5 AAS92974	Aas92974 DNA encod
C 32	41.8	12.7	10614	4 AAS46259	Aas46259 DNA encod
C 33	41.4	12.5	37856	3 AAA11992	Aaa11992 S. cellul
34	41	12.4	396	6 AAL50587	Aal50587 Bovine ad
35	40.6	12.3	354	6 ABN75809	Abn75809 Human ORF
36	40.2	12.2	3256	5 AAS92977	Aas92977 DNA encod
37	39.6	12.0	3807	7 ACA38183	Aca38183 Prokaryot
38	39.2	11.9	572	2 AAQ3718	Aaq3718 Mel clust
39	39.2	11.9	1442	2 AAQ92327	Aaq92327 Chloropla
40	39.2	11.9	1883	3 AAA28414	Aaa28414 ORF-438-I
41	39.2	11.9	4294	2 AAQ92325	Aaq92325 Plasmid p
C 42	38.8	11.8	2000	7 ADA71938	Ada71938 Rice gene
C 43	38.6	11.7	113193	7 AAD54645	Aad54645 Streptomy
C 44	38.4	11.6	110000	2 AAT42063_03	Continuation (4 of
45	38	11.5	426	7 ABZ40263	Abz40263 N. gonorr

ALIGNMENTS

RESULT 1
ACC48896
ID ACC48896 standard; DNA; 8766 BP.

XX ACC48896;

DT 11-AUG-2003 (first entry)

DE Rhodococcus ruber eth gene cluster involved in ETBE degradation.
XX Cytochrome P450; ethA; ethB; ethC; ethD; ethR; fuel; ether;
KW ethyl tert-butyl ether; ETBE; degradation; bioremediation;
KW soil decontamination; pollutant; biosensor; gene; ds.
XX Rhodococcus ruber.

XX Key Location/Qualifiers
FT CDS 461..4858
FT /*tag= a

FT /product= "EthR"
FT /note= "the CDS is claimed in Claim 2, encodes ABR42004"
FT CDS 1640..2878
FT /*tag= b

FT /product= "EthA"
FT /note= "the CDS is claimed in Claim 2, encodes ABR42000"
FT CDS 2948..4150
FT /*tag= c

FT /product= "EthB"
FT /note= "the CDS is claimed in Claim 2, encodes ABR42001"
FT CDS 4156..4476
FT /*tag= d

FT /product= "EthC"
FT /note= "the CDS is claimed in Claim 2, encodes ABR42002"
FT CDS 4547..4858
FT /*tag= e

FT /product= "EthD"
FT /note= "the CDS is claimed in Claim 2, encodes ABR42003"
FT CDS 4547..4858
FT /*tag= e

EP1270722-A1.

02-JAN-2003.

22-JUN-2001; 2001EP-00401667.

22-JUN-2001; 2001EP-00401667.

XX

PA	(INSP) INST PASTEUR.
PA	(INSF) INST FRANCAIS DU PETROLE.
XX	
PI	Chauvaux S, Miras I, Beguin P;
XX	
DR	WPI: 2003-334761/32.
XX	
DR	P-PSDB; ABR42000, ABR42001, ABR42002, ABR42003, ABR42004.
XX	
PT	New Rhodococcus ruber nucleic acid, useful for transforming bacteria for
PT	depolluting soil contaminated with ethyl tert-butyl ether comprises the
PT	cytochrome p-450 gene cluster involved in cleavage of ethyl tert-butyl
PT	ether.
XX	
PS	Claim 7; Page 25-27; 58pp; English.
XX	
CC	The present sequence comprises the eth gene cluster of Rhodococcus ruber
CC	strain CNCM I-1889, which encodes a cytochrome P450 system involved in
CC	the cleavage of ether fuel additions. Expression of this gene cluster is
CC	induced by the fuel additive ethyl tert-butyl ether (ETBE). Loss of
CC	ability to degrade ETBE is associated with a chromosomal rearrangement
CC	resulting in a 14.3 kb deletion including the eth gene cluster. This
CC	rearrangement occurs between two 5.6 kb duplicated sequences. The eth
CC	gene cluster comprises the ethA, ethB, ethC, ethD and ethR open reading
CC	frames. These encode a transcriptional activator (EthR), a ferredoxin
CC	reductase (EthA), an ETBE-inducible cytochrome P450 (EthB), a ferredoxin
CC	(EthC) and an ETBE-inducible unknown protein (EthN). The identification
CC	of the eth gene cluster allows the construction of probes and biosensors
CC	for detecting pollution by an ether fuel and for assessing the potential
CC	of a contaminated soil or effluent to cleave the ether fuel. Also
CC	provided are recombinant bacteria, plant, fungus or yeast cells
CC	comprising a vector encoding the eth genes and which are capable of ETBE
CC	(and tert-butyl alcohol) degradation or mineralization, for use in ether
CC	fuel bioremediation of a contaminated soil or effluent
XX	
SQ	Sequence 8766 BP; 1433 A; 2677 C; 3150 G; 1498 T; 0 U; 0 Other;
	Query Match 22.4%; Score 74; DB 7; Length 8766;
	Best Local Similarity 54.0%; Pred. No. 5.8e-10;
	Matches 174; Conservative 0; Mismatches 145; Indels 3; Gaps 1;
QY	2 TGACCGCAAAGGTCCGGTGATCTTCGCGCAGCGCGGCTTCGAGCATCTGTCGAAA 61
Db	4148 TGATGCCGATGCCCAAGATCACCTTCTCCCAATCGGCGGTCTGCATCACCGTCGATG 4207
QY	62 CCGAAGCGGGAGTAGCTCTCATGAAGCGCGCTTCTGAACGGCGTGACGGTATCGAAG 121
Db	4208 CGTCCGTGGACCAGAGCGTCATGACGGCGCGCTGCGCGCAGGTATCGACGGCATCTCG 4267
QY	122 CCGTTTCGGGGGGCGCTGTGCCTCGCGCAGCTGCCAGTTTACGTTGGCCCCGAGTGGC 181
Db	4268 CCGAGTSCGGTGGCAAGCCACGCTGTTCCACTCGCAGCTGACTCGAACCGAGCAAC 4327
QY	182 TAGATCGCTGAACCCGCGAGTGAGACCGAAGAAATCTCGATTG---CGTAGCGG 238
Db	4328 TCGGGCTCTTCGGGACCTGAGCGCCGGAAGAAGACGACATCTCGACTTCGCGCAGGCCG 4387
QY	239 AACGTGGCGCCGANTTCGGGGCTGTCTTCGCCAGATCCGCGCTTACCAGACCTGCTCGACGCC 298
Db	4388 AACGTCGAGCAACAGCGCGTGGGTGTGAGTCCGGTCAACCCAGACCTCGACGCCG 4447
QY	299 TGACCTTGGAACTGCGGAAGGC 320
Db	4448 TCCGGCTGGAGTCCCCGACGC 4469
RESULT 2	
ACC48895	
ID	ACC48895 standard; DNA; 12342 BP.
XX	
AC	ACC48895;
XX	
DT	11-AUG-2003 (first entry)

DE	Rhodococcus ruber eth gene cluster deleted region.		
XX			
KW	Cytochrome P450; ethA; ethB; ethC; ethD; ethR; fuel; ether;		
KW	ethyl tert-butyl ether; ETBE; degradation; bioremediation;		
KW	soil decontamination; pollutant; biosensor; gene; ds.		
XX			
OS	Rhodococcus ruber.		
XX			
FH	Location/Qualifiers		
FT	CDS	461..4858	
FT		/*tag= a	
FT		/product= "EthR"	
FT		/note= "the CDS is claimed in Claim 2, encodes ABR42004"	
FT	CDS	1640..2878	
FT		/*tag= b	
FT		/product= "ethA"	
FT		/note= "the CDS is claimed in Claim 2, encodes ABR42000"	
FT	CDS	2948..4150	
FT		/*tag= c	
FT		/product= "EthB"	
FT		/note= "the CDS is claimed in Claim 2, encodes ABR42001"	
FT	CDS	4156..4476	
FT		/*tag= d	
FT		/product= "EthC"	
FT		/note= "the CDS is claimed in Claim 2, encodes ABR42002"	
FT	CDS	4547..4858	
FT		/*tag= e	
FT		/product= "EthD"	
FT		/note= "the CDS is claimed in Claim 2, encodes ABR42003"	
XX			
EN	EP1270722-A1.		
XX			
PD	02-JAN-2003.		
XX			
XX	22-JUN-2001; 2001EP-00401667.		
XX			
PR	22-JUN-2001; 2001EP-00401667.		
XX			
PA	(INSP) INST PASTEUR.		
PA	(INSP) INST FRANCAIS DU PETROLE.		
XX			
PI	Chauvaux S, Miras I, Beguin P;		
XX			
DR	WPI; 2003-334761/32.		
DR	P-FSDB; ABR42000, ABR42001, ABR42002, ABR42003, ABR42004.		
XX			
PT	New Rhodococcus ruber nucleic acid, useful for transforming bacteria for		
PT	depolluting soil contaminated with ethyl tert-butyl ether comprises the		
PT	cytochrome P-450 gene cluster involved in cleavage of ethyl tert-butyl		
PT	ether.		
XX			
PS	Claim 3; Page 22-25; 58pp; English.		
XX			
CC	The present sequence is a region of Rhodococcus ruber strain CNCM I-1889		
CC	DNA that is deleted in strains that have lost the ability to cleave the		
CC	ether fuel additive ethyl tert-butyl ether (ETBE). Loss of ability to		
CC	degrade ETBE is associated with a chromosomal rearrangement resulting in		
CC	a 14.3 kb deletion including the eth gene cluster. This rearrangement		
CC	occurs between two 5.6 kb duplicated sequences. The eth gene cluster		
CC	comprises the ethA, ethB, ethC, ethD and ethR open reading frames. These		
CC	encode a transcriptional activator (EthR), a ferredoxin reductase (EthA)		
CC	an ETBE-inducible cytochrome P450 (EthB), a ferredoxin (EthC) and an ETB		
CC	-inducible unknown protein (EthD). The identification of the eth gene		
CC	cluster allows the construction of probes and biosensors for detecting		
CC	pollution by an ether fuel and for assessing the potential of a		
CC	contaminated soil or effluent to cleave the ether fuel. Also provided ar		
CC	recombinant bacteria, plant, fungus or yeast cells comprising a vector		
CC	encoding the eth genes and which are capable of ETBE (and tert-butyl		
CC	alcohol) degradation or mineralization, for use in ether fuel		
CC	bioremediation of a contaminated soil or effluent		
XX			
SQ	Sequence 12342 BP; 2065 A; 3752 C; 4311 G; 2214 T; 0 U; 0 Other;		

RESULT 2

ACC48895

ID ACC48895 standard; DNA; 12342 BP.

XX

AC ACC48895;

XX
DE
11-210-200 (fig)

DT 11-AUG-2003 (first entry)
YY

4


```
Query Match      22.4%; Score 74; DB 7; Length 12342;
Best Local Similarity 54.0%; Pred. No. 6.2e-10;
Matches 174; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

QY 2 TGACCGCAAGGTCGCGTGTATCTTCGCGCAGCGCGGCTTCGAGCATCTGGTCGAAA 61
Db 4148 TGATGCGGATGCCCAAGATCACCTTCTCCAAATCGGACGGTGTGATCACCGTGTGATG 4207

QY 62 CCGAGCGGGAGATGCTCATGGAAGCGGCGCTTCGAACGGGTGGACGGTATCGAAG 121
Db 4208 CGTCTGTCGACGAGCGTATGACGAGCGCGCTTCGCGCAGGTATCGACGGCATCTCTCG 4267

QY 122 CCGTTTCGCGGGCGCTGTGCTTCCGCGCAGCGCGGCTTCGAGCATCTGGTCGAAA 181
Db 4268 CCGAGTCGCTGTGCAACCGCACTGTTCACCTGCCACGCTGACGTCGAACCGGAGCAAC 4327

QY 182 TAGATGCGCTGAACCGCGGAGTGAGACCGAAGACGAAATGCTCGATTG---CGTAGCGG 238
Db 4328 TCGGGCTCTCTCGCGGACCTGAGCGCGCGAAGACGACATGCTCGACTGCGCGAGCGCG 4387

QY 239 AAGTGGCGCGCATTCGCGGCTGTCTCGCAGATCGCCCTTACCGACCTGCTCGACGGCG 298
Db 4388 AACGTCGAGCAACAGCGCGCTGCGGTGTGAGTCGCGGTACACCGACCTCGACGGCG 4447

QY 299 TGACCCCTGGAACCTGCCGAAGCGC 320
Db 4448 TCGGCTGAGGTCCTCCGACGCGC 4469

RESULT 3
ACC48894
ID ACC48894 standard; DNA; 23656 BP.
AC ACC48894;
XX
XX 11-AUG-2003 (first entry)
XX Rhodococcus ruber eth gene region involved in ETBE degradation.
XX Cytochrome P450; ethA; ethB; ethC; ethD; ethR; fuel; ether;
XX ethyl tert-butyl ether; ETBE; degradation; bioremediation;
XX soil decontamination; pollutant; biosensor; gene; ds.
XX Rhodococcus ruber.
XX
XX Key Location/Qualifiers
FH repeat_region 3519..23436
FT /*tag= c
FT /*type= FLANKING
FT repeat_unit 3519..9094
FT /*tag= d
FT /*note= "duplicated sequence, referred to in Claim 4"
FT misc_feature 9095..21436
FT /*tag= a
FT /*note= "region deleted in ETBE-negative bacteria,
FT referred in Claim 3"
FT misc_feature 9095..17860
FT /*tag= b
FT /*note= "eth gene cluster region, referred to in Claim 7"
FT CDS 9555..13952
FT /*tag= f
FT /*product= "EthR"
FT 10734..11972
FT CDS /*tag= g
FT /*product= "EthA"
FT /*note= "the CDS is claimed in Claim 2, encodes ABR42004"
FT 12042..13244
FT CDS /*tag= h
FT /*product= "EthB"
FT /*note= "the CDS is claimed in Claim 2, encodes ABR42001"
FT 13250..13570
FT CDS /*tag= i
FT 13570..13570
```

```
FT FT /product= "EthC"
FT FT /note= "the CDS is claimed in Claim 2, encodes ABR42002"
FT FT 13641..13952
FT FT /*tag= j
FT FT /product= "EthD"
FT FT /note= "the CDS is claimed in Claim 2, encodes ABR42003"
FT FT 17861..23536
FT FT /*tag= e
FT FT /note= "duplicated sequence, referred to in Claim 4"

PN EP1270722-A1.
XX
XX 02-JAN-2003.
XX
XX 22-JUN-2001; 2001EP-00401667.
XX
XX 22-JUN-2001; 2001EP-00401667.
XX (INSP ) INST PASTEUR.
XX (INSF ) INST FRANCAIS DU PETROLE.
XX Chauvaux S, Miras I, Beguin P;
XX
XX WPI; 2003-334761/32.
XX P-PSDB; ABR42000, ABR42001, ABR42002, ABR42003, ABR42004.
XX
XX New Rhodococcus ruber nucleic acid, useful for transforming bacteria for
XX depolluting soil contaminated with ethyl tert-butyl ether comprises the
XX cytochrome P-450 gene cluster involved in cleavage of ethyl tert-butyl
XX ether.
XX
XX Claim 7; Page 15-21; 58pp; English.
XX
XX The present sequence is a BamHI fragment of Rhodococcus ruber strain CNCM
XX I-1889 DNA that includes an eth gene cluster encoding a cytochrome P450
XX system involved in the cleavage of ether fuel additives. Expression of
XX the gene cluster is induced by ethyl tert-butyl ether (ETBE). Loss of
XX ability to degrade ETBE is associated with a chromosomal rearrangement
XX resulting in a 14.3 kb deletion including the eth gene cluster. This
XX rearrangement occurs between two 5.6 kb duplicated sequences. The eth
XX gene cluster comprises the ethA, ethB, ethC, ethD and ethR open reading
XX frames. These encode a transcriptional activator (EthR), a ferredoxin
XX reductase (EthA), an ETBE-inducible cytochrome P450 (EthB), a ferredoxin
XX (EthC) and an ETBE-inducible unknown protein (EthD). Identification of
XX the eth gene cluster allows the construction of probes and biosensors for
XX detecting pollution by an ether fuel and for assessing the potential of a
XX contaminated soil or effluent to cleave the ether fuel. Also provided are
XX recombinant bacteria, plant, fungus or yeast cells comprising a vector
XX encoding the eth genes and which are capable of ETBE (and tert-butyl
XX alcohol) degradation or mineralization, for use in ether fuel
XX bioremediation of a contaminated soil or effluent
XX
XX Sequence 23656 BP; 3867 A; 7590 C; 7936 G; 4263 T; 0 U; 0 Other;
```

```
Query Match      22.4%; Score 74; DB 7; Length 23656;
Best Local Similarity 54.0%; Pred. No. 6.9e-10;
Matches 174; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

QY 2 TGACCGCAAGGTCGCGTGTATCTTCGCGCAGCGCGGCTTCGAGCATCTGGTCGAAA 61
Db 13242 TGATGCGGATGCCCAAGATCACCTTCTCCAAATCGGACGGTGTGATCACCGTGTGATG 13301

QY 62 CCGAGCGGGAGTATCGCTTCATGGAAGCGCGCGCTTCGAAAGCGGTGGACGGTATCGAAG 121
Db 13302 CGTCTGTGACCGAGAGCGTTCATGCGCGCGCGCTTCGCGCAGGTATCGACGGCATCTCG 13361

QY 122 CCGTTTCGCGGGCGCTGTGCTTCCGCGCAGCGCGGCTTCGAGCATCTGGTCGAAA 181
Db 13362 CCGAGTCGCTGTGCAACCGCACTGTTCACCTGCCACGCTGACGTCGAACCGGAGCAAC 13421

QY 182 TAGATGCGCTGAACCGCGGAGTGAGACCGAAGACGAAATGCTCGATTG---CGTAGCGG 238
Db 13422 TCGGGCTCTCTCGCGACCTGAGCGCGGAGAGACGACATGCTCGACTGCGCGGAGCGCG 13481
```

QY 239 AAGTGGCGGCGATTGCGGGTGTCTGCGAGATCGGCTTACGACCTGCTCGACGCC 298
|||||
Db 13482 AACCTCGGAGACACACCGCGCTGGCGTGTACGTGCGGGTACCAACCTCGACGGC 13541
|||||
QY 299 TGACCTGGAAGTGGCGAAGC 320
|||||
Db 13542 TCGGCTGGAGGTCCCGGACG 13563
|||||

RESULT 4
AAH65575/c
ID AAH65575 standard; DNA; 306 BP.

XX AAH65575;
AC
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 610.
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-00127688.
XX
PR 16-DEC-1999; 99JP-00377484.
PR 07-APR-2000; 2000JP-00159162.
PR 03-AUG-2000; 2000JP-00280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
DR WPI; 2001-376931/40.
DR P-PSDB; AAG90356.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX
PS Claim 8; SEQ ID NO 610; 246pp + Sequence Listing; English.
XX

CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and analysing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC Coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the European Patent Office
XX

SQ Sequence 306 BP; 75 A; 92 C; 71 G; 68 T; 0 U; 0 Other;

Query Match 19.9%; Score 65.6; DB 5; Length 306;
Best Local Similarity 58.2%; Pred. No. 5.7e-08;
Matches 135; Conservative 0; Mismatches 94; Indels 3; Gaps 1;

QY 56 TCGAACCAGCGGAGATCGCTCATGAGCGCGCGCTTGTGACGCGTGGACGCTA 115
|||||
Db 286 TCGAGCGGACTGTTGGTGATTGATGTAAGACCGCAGTCCGAACGGAGTGCTGGAA 227
|||||
QY 116 TCGAAGCCGTTTTCGGGGGGCGCGTGTGCTGCGCCACGTCGCTTACGTTGCGCCG 175
|||||

Db 226 TTGTTGCTGAATGGCGGGTTCCTTATCGTGTGCAACCTGCCATGTGTTGTGACCTG 167
|||||
QY 176 AGTGGCTAGATGCGCTGAAACCCCGAGTGAGACCGAAGACGAATGCT---CGATTGG 232
|||||
Db 166 CACAGTATGATGCGCTTCCCCCAATGGAGAGATGAAGATGCTGTGGGGTGCTG 107
|||||
QY 233 TAGCGGAACGTGGCGCGCATTCGCGGTGTCTGCGCAGATCCGCCCTTACCGA 284
|||||
Db 106 CCGTGGACCGTGAGGATTGCTCCGTTGTCTTGCCCAATCAAGTGCACCGA 55
|||||

RESULT 5
AAH65577
ID AAH65577 standard; DNA; 318 BP.

XX AAH65577;
AC
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 612.
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-00127688.
XX
PR 16-DEC-1999; 99JP-00377484.
PR 07-APR-2000; 2000JP-00159162.
PR 03-AUG-2000; 2000JP-00280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
DR WPI; 2001-376931/40.
DR P-PSDB; AAG90358.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX

PS Claim 8; SEQ ID NO 612; 246pp + Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and analysing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC Coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the European Patent Office
XX

SQ Sequence 318 BP; 71 A; 75 C; 90 G; 82 T; 0 U; 0 Other;

Query Match 19.9%; Score 65.6; DB 5; Length 318;
Best Local Similarity 58.2%; Pred. No. 5.7e-08;
Matches 135; Conservative 0; Mismatches 94; Indels 3; Gaps 1;

QY 56 TCGAACCAGCGGAGATCGCTCATGAGCGCGCGCTTGTGACGCGTGGACGCTA 115
|||||
Db 47 TCGAGGCGACTGTTGGTGATTGATGTAAGACCGCAGTCCGAACGGAGTGCTGGAA 106
|||||

QY 116 TCGAAGCGTTTCGGGGGCGCTGTGCTGCGCCACGTCGACGTTTACGTTGGCCCG 175
Db 107 TTGTTGCTGAATGCGGCGTTCTTATCGTGTGCAACTGCGATGTTGTGACCCG 166
QY 176 AGTGGCTAGATGCTGCTGAACCGCGAGTGAGACCGAAGACGAATGCT---CGATTGG 232
Db 167 CACAGTATGATGGCTTCCCGCAATGAGGAGATGGAAGATGAATGCTGTGGGTGCTG 226
QY 233 TAGCGGAACGTGCGCGCATTCGCGCTGTCCTGCCAGATCCGCCCTTACCGA 284
Db 227 CCGTGGACCGTGAGGATGCTCCCGTTGTCTTGCCCAATCAAGTCAACCGA 278

RESULT 6

AAAF71697

ID AAF71697 standard; DNA; 441 BP.

XX AAF71697;

AC AAF71697;

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:675.

XX Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study; ds.

XX Corynebacterium glutamicum.

XX W0200100844-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB000943.

XX 25-JUN-1999; 99US-0141031P.

XX 08-JUL-1999; 99DE-01031412.

XX 08-JUL-1999; 99DE-01031413.

XX 08-JUL-1999; 99DE-01031419.

XX 08-JUL-1999; 99DE-01031420.

XX 08-JUL-1999; 99DE-01031424.

XX 08-JUL-1999; 99DE-01031428.

XX 08-JUL-1999; 99DE-01031431.

XX 08-JUL-1999; 99DE-01031433.

XX 08-JUL-1999; 99DE-01031434.

XX 08-JUL-1999; 99DE-01031510.

XX 08-JUL-1999; 99DE-01031562.

XX 08-JUL-1999; 99DE-01031634.

XX 09-JUL-1999; 99DE-01032180.

XX 09-JUL-1999; 99DE-01032227.

XX 09-JUL-1999; 99DE-01032230.

XX 09-JUL-1999; 99US-0143208P.

XX 14-JUL-1999; 99DE-01032924.

XX 14-JUL-1999; 99DE-01032973.

XX 14-JUL-1999; 99DE-01033005.

XX 27-AUG-1999; 99DE-01040765.

XX 31-AUG-1999; 99US-0151572P.

XX 03-SEP-1999; 99DE-01042076.

XX 03-SEP-1999; 99DE-01042079.

XX 03-SEP-1999; 99DE-01042086.

XX 03-SEP-1999; 99DE-01042087.

XX 03-SEP-1999; 99DE-01042088.

XX 03-SEP-1999; 99DE-01042095.

XX 03-SEP-1999; 99DE-01042123.

XX 03-SEP-1999; 99DE-01042125.

XX (BADI) BASF AG.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX WPI; 2001-061975/07.

DR P-PSDB; AAB79580.

XX

PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes.

XX Claim 3; Page 1096; 1246pp; English.

XX

CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243
CC to AAB 79633 which are involved in carbon metabolism and energy
CC production. The C. glutamicum SMP gene can be used in vectors (II) for
CC expression in host cells and production or modulation of production of
CC fine chemicals, such as, an organic acid, a proteinogenic or
CC nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a
CC nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid,
CC a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a
CC polyketide, or an enzyme. The presence of (I) or SMP proteins (III)
CC encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to C.
CC glutamicum, identify and localise C. glutamicum sequences of interest, in
CC evolutionary studies, in determining SMP protein regions required for
CC function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH)

XX Sequence 441 BP; 109 A; 102 C; 116 G; 114 T; 0 U; 0 Other;

SQ

Query Match 19.9%; Score 65.6; DB 4; Length 441;

Best Local Similarity 58.2%; Pred. No. 6.1e-08;

Matches 135; Conservative 0; Mismatches 94; Indels 3; Gaps 1;

QY 56 TCGAACCAGACGGGAGTATCGCTCATGGAAGCGCGCTTCTGAACGGCTGGACGGTA 115

Db 147 TCGAGCGACTGTTGGTCAATCAGTATGAGACCGCAGTCCGAACGGATGCTGAA 206

QY 116 TCGAAGCGCTTTGGGGGGCGCTGTGCTGCGGCACGTGCCACGTTTACGTTGGCCCCG 175

Db 207 TTGTTGCTGAATGCGGGCGTTCCTTATCGTGTGCAACCTGCATGTTGTTGACCTG 266

QY 176 AGTGGCTAGATGCGCTGAAACCGCGAGTGAGACCGAAGACGAATGCT---CGATTGCG 232

Db 267 CACAGTATGATGCGCTTCCCGCAATGAGAGATGAGATGAATGCTGTGGGGTCTG 326

QY 233 TAGCGGAACGTGCGCGCATTCGCGGCTGTCTGCCAGATCCGCCCTTACCGA 284

Db 327 CCGTGGACCGTGAGGATGCTCCCGTTGTCTTGCCCAATCAAGTCAACCGA 378

RESULT 7

AAH68525/c

ID AAH68525 standard; DNA; 349980 BP.

XX AAH68525;

AC AAH68525;

XX 26-SEP-2001 (first entry)

DT

XX C glutamicum coding sequence fragment SEQ ID NO: 7060.

DE

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

XX

OS Corynebacterium glutamicum.

XX

EP1108790-A2.

XX

20-JUN-2001.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PF 18-DEC-2000; 2000EP-00127688.
XX
XX
PR 16-DEC-1999; 99JP-00377484.
PR 07-APR-2000; 2000JP-00159182.
PR 03-AUG-2000; 2000JP-00280986.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX
XX Disclosure; SEQ ID NO 7060; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and analyzing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids, described
CC particularly L-lysine. The present sequence is a nucleic acid, described
CC in the exemplification of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the European Patent Office
XX
SQ Sequence 349980 BP; 80289 A; 91081 C; 97378 G; 81232 T; 0 U; 0 Other;

Query Match 19.9%; Score 65.6; DB 5; Length 349980;
Best Local Similarity 58.2%; Pred. No. 1.9e-07;
Matches 135; Conservative 0; Mismatches 94; Indels 3; Gaps 1;

QY 56 TCGAAACCGAAGCGGAGTATCGCTCATGGAAGCGCGCGTTCTGAACGCGCTGACGCGTA 115
DB 262917 TCGAGGCGACTGTGGTGATTTCAGTAATGGAGACCGCAGTCGAAACGGAGTGCTGAA 262858

QY 116 TCGAAGCCGTTTGGGGGCGCTGTGCTCGCGCCACGTGCGCACGTTTACGTTGCCCCG 175
DB 262857 TTGTTGCTGAATGCGGCGGTTCCCTTATCGTGTGCAACCTGCCATGTGTTGTGACCCGTG 262798

QY 176 AGTGGCTAGATCGCTGAAACCGCGGAGTGAGACCGAAGACGAAATGCT---CGATTGCG 232
DB 262797 CACAGTATGATGGCTTCCCCCAATGGAGAGATGGAAGATGAATGTTGGGGTGCTG 262738

QY 233 TAGCGAAGCGTGGCGCGAATTCGGGCTGTCTCTGCGAGATCGGCTTTACCGA 284
DB 262737 CGGTGACCGTGAGGATTGCTCCCGTTTCTTGTGCAAAATCAAGGTACCGA 262686

RESULT 8
ABLO7331
ID ABL07331 standard; cDNA; 389 BP.
XX
XX ABL07331;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 16475.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD

XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR P-PSDB; ABB63228.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 16475; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fip.wipo.int/pub/published_pct_sequences
XX
XX Sequence 389 BP; 106 A; 97 C; 107 G; 79 T; 0 U; 0 Other;
SQ

Query Match 19.0%; Score 62.8; DB 4; Length 389;
Best Local Similarity 56.2%; Pred. No. 3.3e-07;
Matches 119; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 111 CGTATCGAAGCCGTTTGGGGGCGCCCTGTGCTCGCCAGCGTCCACGTTTACGTTGG 170
DB 97 CGAGATGGAAGGCGCCTGTGAGGCTTCGCTGGCCTGCACCCACCTGTACGTTGACGTCCA 156

QY 171 CCCGAGTGGCTAGATCGCTGAAACCGCGAGTGAGACCGAAGACGAAATGCTCGATTG 230
DB 157 GCATGATTACCTGCAGAGTTAAAGAGCCGAGGAGCAGGAGGAGCTGCTGATAT 216

QY 231 CGTAGCGGAACGTGCGCCGCAATTCGGGCTGTCTCCAGATCCGCTTACCGACCTGCT 290
DB 217 GGCGCCATTTCTCGCGGAGAACTCCCGGCTCGGCTGTGAGTACTCTCGACAAGAGTAT 276

QY 291 CGAGCGCTGACCCCTGGAATCGCCGAGGC 320
DB 277 GGAGGGCATGGAACTGGAGCTGCCAAGGC 306

RESULT 9
ABLO7330
ID ABL07330 standard; cDNA; 2389 BP.
XX
XX ABL07330;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 16472.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US009231.
PF

```
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX P-PSDB; ABB63227.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 16472; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2389 BP; 702 A; 510 C; 492 G; 685 T; 0 U; 0 Other;
SQ
Query Match 19.0%; Score 62.8; DB 4; Length 2389;
Best Local Similarity 56.2%; Pred. No. 4.6e-07;
Matches 118; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 111 CGGTATCGAAGCCGTTTCCGGGGCGCTGTGCTGGCCGACGCTGACGTTTACCTGG 170
DB 1097 CGAGATGGAAGGCCCTGTGAGGCTTCGCTGGCTGCACCACTGTACGTGACGTCCA 1156
QY 171 CCCGAGTGGCTAGATCGCTGAAACCCGCGAGTGAGACCGAAGACGAAATGCTCGATTG 230
DB 1157 GCATGATTACCTGCAGAAATTAAAGAGGCGGAGGAGGAGGAGCGACCTGCTGGATAT 1216
QY 231 CGTAGCGGAGCGTGGCGGATTCGCGCTGTCTGCGCAGATCGGCTTACCGACCTGCT 290
DB 1217 GCGGCCATTTCTGCGCGAGAACTCCCGGCTCGGCTGTGATGATCTCTGACAAGAGTAT 1276
QY 291 CGACGGCTGACCTGGAACCTGCCAAGGC 320
DB 1277 GGAGGGCATGGAACCTGGAGCTGCCAAGGC 1306
RESULT 10
ABL07682
XX ABL07682 standard; cDNA; 4531 BP.
XX
XX ABL07682;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 17528.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 23-MAR-2000; 2000US-0191637P.
```

```
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX P-PSDB; ABB63579.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 17528; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 4531 BP; 1185 A; 994 C; 1043 G; 1309 T; 0 U; 0 Other;
SQ
Query Match 19.0%; Score 62.8; DB 4; Length 4531;
Best Local Similarity 56.2%; Pred. No. 5.1e-07;
Matches 118; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 111 CGGTATCGAAGCCGTTTCCGGGGCGCTGTGCTGGCCGACGCTGACGTTTACGTTGG 170
DB 339 CGAGATGGAAGGCCCTGTGAGGCTTCGCTGGCTGCACCACTGTACGTTAGCTCCA 398
QY 171 CCCGAGTGGCTAGATCGCTGAAACCCGCGAGTGAGACCGAAGACGAAATGCTCGATTG 230
DB 399 GCATGATTACCTGCAGAAATTAAAGAGGCGGAGGAGGAGGAGGAGCGACCTGCTGGATAT 458
QY 231 CGTAGCGGAGCGTGGCGGATTCGCGCTGTCTGCGCAGATCGGCTTACCGACCTGCT 290
DB 459 GCGGCCATTTCTGCGCGAGAACTCCCGGCTCGGCTGTGATGATCTCTCGACAGAGTAT 518
QY 291 CGACGGCTGACCTGGAACCTGCCAAGGC 320
DB 519 GGAGGGCATGGAACCTGGAGCTGCCAAGGC 548
RESULT 11
ABL07688
XX ABL07688 standard; cDNA; 4531 BP.
XX
XX ABL07688;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 17546.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
```

PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-556860/75.
DR P-PSDB; ABB63585.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 17546; 21pp + Sequence Listing; English.
PS
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4531 BP; 1185 A; 994 C; 1043 G; 1309 T; 0 U; 0 Other;
Query Match 19.0%; Score 62.8; DB 4; Length 4531;
Best Local Similarity 56.2%; Pred. No. 5.1e-07;
Matches 118; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 111 CGGTATCGAAGCGTTTGGGGGGCGCTGTCCTGCGCCACGTCGCCACGTTTACGTGG 170
DB 339 CGAGATGGAGGGCGCTGTGAGGCTGCTGCGCTGCACCCACTGTACGTGTACGTCCA 398
QY 171 CCCCAGTGGCTAGATGCGCTGAAACCGCCGAGTGAGACCGAAGAAATGCTCGATTG 230
DB 399 GCATGATTACCTGCAGAGCTTAAAGAGCGCCGAGGAGGAGGACGACCTGCTGGATAT 458
QY 231 CGTAGGGAGCGTGGCCCGCATTCGCGGCTGTCCTGCCAGATCCCGCTTACCGACCTGCT 290
DB 459 GGGCGCATTTTGGCGGAGAACTCCCGCTCGGCTGTCTAGATCTCTCGACAAAGAT 518
QY 291 CGACGCGCTGACCTGGAACTCCCGAAGGC 320
DB 519 GGAGGGCATGGAACTGGAGCTGCCCAAGGC 548
RESULT 12
ACH38677
ID ACH38677 standard; cDNA; 458 BP.
XX
XX ACH38677;
XX
DT 13-OCT-2003 (first entry)
DE Human foetal brain cDNA #44.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX
XX Homo sapiens.
XX
XX US2003073623-A1.
PN
PD 17-APR-2003.
XX
XX 30-JUL-2001; 2001US-00918995.
PF
XX
XX 30-JUL-2001; 2001US-00918995.
PR
XX
XX (DRNA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA

PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI; 2003-615964/58.
DR
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 25889; 44pp; English.
PS
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 458 BP; 90 A; 132 C; 152 G; 79 T; 0 U; 5 Other;
Query Match 18.3%; Score 60.4; DB 8; Length 458;
Best Local Similarity 55.1%; Pred. No. 1.5e-06;
Matches 118; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 105 CFTGGAGGGTATCGAAGCGCTTTTGGGGGGCGCTGTCGCTGGCCACGTCGCCACGTTTA 164
DB 211 CGGGTGGAGCTTGAAGGGGCTGTGAAGCTCTCTGGCTGTCTCCACTGCTGCTGTA 270
QY 165 CGTTGGCCCCGAGTGGCTAGATCGCTGAAACCGCCGAGTGAGACCGAAGACGAATGCT 224
DB 271 TGTGAGTGAAGACCACTGGATCTCTGCTCTCTCCGAGGAGGAGGAGACGACATGCT 330
QY 225 CGATTGGCTAGCGAAGCGTGGCGCGCTGTCGCGCTGTCCTGCCAGATCCGCTTACCGA 284
DB 331 AGACATGGCCCCCTCTCTCCAGAGAACTCGCGGCTGGGCTGCCAGATTGCTGACACC 390
QY 285 CTTGCTCGACGGCTGACCTGAACTGGAAGTGGCCGAAG 318
DB 391 GGAGCTGGAAGGAGCGGAATTCACCTGCTGCCAAG 424
RESULT 13
ACH33408
ID ACH33408 standard; cDNA; 473 BP.
XX
XX ACH33408;
XX
DT 13-OCT-2003 (first entry)
DE Human endothelial cell cDNA #1541.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX
XX Homo sapiens.
XX
XX US2003073623-A1.
PN
PD 17-APR-2003.

XX PF 30-JUL-2001; 2001US-00918995.
XX PR 30-JUL-2001; 2001US-00918995.
XX PA (DRMA/) DRMANAC R T.
XX PA (LABA/) LABAT I.
XX PA (STAC/) STACHE-CRAIN B.
XX PA (DICK/) DICKSON M C.
XX PA (JONE/) JONES L W.
XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX DR WPI; 2003-615964/58.
XX PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX PS Claim 1; SEQ ID NO 20620; 44pp; English.
XX CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX SQ Sequence 473 BP; 96 A; 146 C; 146 G; 83 T; 0 U; 2 Other;
Query Match 18.3%; Score 60.4; DB 8; Length 473;
Best Local Similarity 55.1%; Pred. No. 1.5e-06;
Matches 118; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 105 CGTGGACGGTATCGAAGCCGTTTTCGGGGCGCGCTGTGCTGCGCCACGTCGACGTTTA 164
DB 151 CGGGGTGGACCTGGAAGGGGCGCTGTGAAGCCTCCCTGGCTGCTCCACCTGCCATGTGTA 210
QY 165 CGTTGGCCCCGAGTGTAGATCGCTGAAACCGCGAGTGAGACGAGAGAAATGCT 224
DB 211 TGTGAGTGAAGACCACTGGATCTCTGCTCTCTCCGAGGAGAGGAGACGATGCT 270
QY 225 CGATTGCGTAGCGGAACGTGCGCGCATTCGGGGCTGTCTGCGAGATCCGCTTACCGA 284
DB 271 AGACATGGCCCCCTCTCCAGAGAACTCGGGCTGGCTGCCAGATTGCTGCACACC 330
QY 285 CTGTCTGACGGCTGACCTTGGAACTGCCGAAG 318
DB 331 GGAGCTGGAAGGAGCGGAATTACCCCTGCCCAAG 364
RESULT 14
ABQ60909
ID ABQ60909 standard; cDNA; 834 BP.
XX AC ABQ60909;
XX DT 26-FEB-2003 (first entry)
XX DE Signal peptide HSP-132 encoding sequence.

KW Neuroprotective; immunomodulator; cancer; chromosome 1p; cytostatic;
KW anti-inflammatory; gene therapy; nutritional supplement; wound; burn;
KW ulcer; Alzheimer's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW vulnery; gene; ss.
XX OS Homo sapiens.
XX WO200231111-A2.
XX PD 18-APR-2002.
XX PF 11-OCT-2001; 2001WO-US027760.
XX PR 12-OCT-2000; 2000US-00687527.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-426278/45.
XX DR N-FSDB; ABP43665.
XX PT New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation.
XX PS Claim 1; SEQ ID # 122; 357pp + Sequence Listing; English.
XX CC The invention relates to 446 newly isolated polynucleotide sequences. The
CC activity of polynucleotides of the invention may be described as,
CC vulnery, neuroprotective, immunomodulator, cytostatic and anti-
CC inflammatory. Compositions comprising nucleic acids of the invention are
CC useful for treating a mammalian subject, or as nutritional sources or
CC supplements. These are useful in gene therapy, particularly for treating
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC inflammatory. The nucleic acids and polypeptides are also useful in
CC diagnostic and research methods. The sequences given in records ABQ60788-
CC ABQ61233 represent polynucleotides of the invention. NOTE: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 834 BP; 178 A; 237 C; 285 G; 134 T; 0 U; 0 Other;
Query Match 18.3%; Score 60.4; DB 6; Length 834;
Best Local Similarity 55.1%; Pred. No. 1.7e-06;
Matches 118; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 105 CGTGGACGGTATCGAAGCCGTTTTCGGGGCGCGCTGTGCTGCGCCACGTCGACGTTTA 164
DB 289 CGGGGTGGACCTGGAAGGGGCGCTGTGAAGCCTCCCTGGCTGCTCCACCTGCCATGTGTA 348
QY 165 CGTTGGCCCCGAGTGTAGATCGCTGAAACCGCGAGTGAGACGAGAAATGCT 224
DB 349 TGTGAGTGAAGACCACTGGATCTCTGCTCTCTCCGAGGAGAGGAGAACACATGCT 408
QY 225 CGATTGCGTAGCGGAACGTGCGCGCATTCGCGCTGTCTGCGAGATCCGCTTACCGA 284
DB 409 AGACATGGCCCCCTCTCCAGGAGAACTCGCGCTGGGTGCGCAGATTGTGCTGCACACC 468
QY 285 CTGTCTGACGGCTGACCTTGGAACTGCCGAAG 318
DB 469 GGAGCTGGAAGGAGCGGAATTACCCCTGCCCAAG 502
RESULT 15
AAZ98240
ID AAZ98240 standard; cDNA; 840 BP.
XX

AC AAZ98240;
XX
DT 11-MAY-2000 (first entry)
XX
DE Human signal peptide containing protein HSPP-132 cDNA SEQ ID NO:266.
XX
KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; neurotrophic; neuroprotective; cardiovascular; hepatotropic;
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's diseases; ovulatory defect;
KW muscular dystrophy; ss.
XX
OS Homo sapiens.
XX
FN WO200000610-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US014484.
XX
PR 26-JUN-1999; 98US-0090762P.
PR 31-JUL-1998; 98US-0094983P.
PR 01-OCT-1998; 98US-0102686P.
PR 11-DEC-1998; 98US-0112129P.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
PI Bandman O;
XX
XX WPI: 2000-160673/14.
DR P-PSDB; AAY87355.
XX
XX
PT New human signal peptide-containing proteins useful in treatment,
PT prevention and diagnosis of e.g. cancer, inflammation and cardiovascular
PT disease.
XX
PS Claim 9; Page 326; 327pp; English.
XX
CC AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
CC anticancer, anti-inflammatory, antimicrobial, neurotrophic, hepatotropic,
CC neuroprotective, cardiovascular and antiasthmatic activities, and can be
CC used in gene therapy. HSPPs can be used to treat or prevent disorders
CC associated with decreased activity or function of HSPP. Antagonists of
CC HSPP are used to treat or prevent disorders associated with increased
CC activity or function of HSPP. Such diseases include cell proliferation
CC (including cancer), inflammation, cardiovascular, neurological,
CC reproductive or developmental disorders, (e.g. arteriosclerosis,
CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC asthma, Crohn's disease, microbial or other infections, congestive or
CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC nucleic acids can be used for the recombinant production of HSPP, for
CC detecting HSPP in standard hybridisation and amplification assays (for
CC diagnosis and monitoring), in gene therapy, as antisense, triplex-forming
CC or ribozyme therapeutics, for detecting related sequences or genetic
CC variations, and for chromosomal mapping. HSPP are also used to raise
CC specific antibodies (Ab) and to screen for agonists and antagonists
CC (potential therapeutic agents). Ab are used to diagnose, or monitor, HSPP
CC -related diseases (in usual immunoassays), as therapeutic antagonists, in
CC competitive drug screens, and for purification of HSPP from natural
CC sources
XX
SQ Sequence 840 BP; 178 A; 241 C; 287 G; 134 T; 0 U; 0 Other;

Query Match 18.3%; Score 60.4; DB 3; Length 840;
Best Local Similarity 55.1%; Pred. No. 1.7e-06;

	Matches	118;	Conservative	0;	Mismatches	96;	Indels	0;	Gaps	0;
Qy	105	CCTGACGGTATCGAAGCGTTTGGGGGGCGCCTGTGCTGCGCCACGTGCCACGTTTA	164							
Db	298	CGGGGTGGACCTCGAAGGGGGCCTGTGAAGCCTCCCTGCGCTCCACCTGCCATGTGTA	357							
Qy	165	CGTTGCCCCCGAGTGGCTAGATGCGCTGAAACCGCGGAGTGAGACCGAAGCGAAATGCT	224							
Db	358	TGTGAGTGAAGACCCACCTGGATCTCTGCTCTCTCCGAGGAGGAGAGACGATGCT	417							
Qy	225	CGATTGCGTAGCGGAAACGTGCGCCCGCATTTGCGGGTGTCTCTCCAGATCCCGCTTACCGA	284							
Db	418	AGACATGGCCCCCTCTCTCCAGAGAACTCGGGGCTGGGCTCCAGATTGTGCTGACACC	477							
Qy	285	CCTGCTCGAGCGCCTGACCCCTGGAATGCGCGAAG	318							
Db	478	GGAGCTGGAAGGAGCGGAATTCACCTTGCCCAAG	511							

Search completed: June 19, 2004, 04:29:10
Job time : 149.642 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2004, 04:06:17 ; Search time 963.906 Seconds
(without alignments)

10223.528 Million cell updates/sec

Title: US-10-658-691-2

Perfect score: 330

Sequence: 1 atgaccgcaaaaggtccgct.....tgccgaaggcacagtcacga 330

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba.*

2: em_esthum.*

3: em_estlin.*

4: em_estm.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_estl.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: gb_estfun.*

15: em_estom.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	71.4	21.6	615	14 CF851075	CF851075 psMA016x0
2	70.2	21.3	578	28 AQ651128	AQ651128 Sheared D
3	66.4	20.1	917	14 CF241493	CF241493 AGENCOURT
4	65.8	19.9	624	9 AL859568	AL859568 AL859568

5	65.8	19.9	714	13 BX733466	BX733466 BX733466
6	65.8	19.9	762	14 CF343736	CF343736 AGENCOURT
7	65.8	19.9	880	14 CF376289	CF376289 AGENCOURT
8	65.8	19.9	913	14 CF241492	CF241492 AGENCOURT
9	65.6	19.9	525	28 BH195739	BH195739 TC3-33B9.
10	65.6	19.9	667	12 BM643564	BM643564 170006873
11	64.8	19.6	979	14 CF220100	CF220100 AGENCOURT
12	64.4	19.5	483	14 CB405072	CB405072 OSTR036H1
13	64.4	19.5	488	14 CB385967	CB385967 OSTF036H1
14	62.8	18.0	577	12 BI626767	BI626767 RH57819.5
15	61.8	18.7	632	9 AI562398	AI562398 TENS2389
16	61.4	18.6	624	9 AI562390	AI562390 TENS2384
17	60.4	18.3	493	12 BM823886	BM823886 K-EST0095
18	60.4	18.3	500	10 AW248266	AW248266 2819638.5
19	60.4	18.3	588	10 BE222719	BE222719 hu51h11.x
20	60.4	18.3	596	10 BE883887	BE883887 601510295
21	60.4	18.3	599	12 BM765293	BM765293 K-EST0064
22	60.4	18.3	608	12 BG831648	BG831648 602765726
23	60.4	18.3	630	9 AA569348	AA569348 nh89b09.s
24	60.4	18.3	636	10 BF036698	BF036698 601459786
25	60.4	18.3	639	9 AI453531	AI453531 t57a11.x
26	60.4	18.3	644	12 BM682117	BM682117 UI-E-B01-
27	60.4	18.3	648	12 BG775214	BG775214 602650217
28	60.4	18.3	704	12 BI667562	BI667562 603292392
29	60.4	18.3	707	13 BU502531	BU502531 AGENCOURT
30	60.4	18.3	714	14 CA421861	CA421861 UI-H-EG0-
31	60.4	18.3	716	13 BQ957421	BQ957421 AGENCOURT
32	60.4	18.3	743	12 BI823913	BI823913 603039086
33	60.4	18.3	747	10 BE744826	BE744826 601575055
34	60.4	18.3	768	10 BE731453	BE731453 601567736
35	60.4	18.3	768	12 BI762440	BI762440 603048832
36	60.4	18.3	773	14 CB987987	CB987987 AGENCOURT
37	60.4	18.3	824	12 BG775846	BG775846 602650217
38	60.4	18.3	883	12 BM045595	BM045595 603623754
39	60.4	18.3	894	14 CD557260	CD557260 AGENCOURT
40	60.4	18.3	901	13 BQ958408	BQ958408 AGENCOURT
41	60.4	18.3	1201	9 AL524608	AL524608 AL524608
42	60.2	18.2	524	9 AI670909	AI670909 wa06902.x
43	59.4	18.0	1029	13 BU502474	BU502474 AGENCOURT
44	59.2	17.9	282	9 AA341290	AA341290 EST46665
45	59	17.9	521	9 AA824331	AA824331 aj30d11.s

ALIGNMENTS

RESULT 1	CF851075	CF851075	615 bp	mRNA	linear	EST 30-OCT-2003
LOCUS	psMA016x004f	USDA-IPAFS:Expression of Phytophthora sojae genes during infection and propagation_sMA	Phytophthora sojae cDNA clone			
DEFINITION	sMA016004 5, mRNA sequence.					
ACCESSION	CF851075	GI:38066729				
VERSION	EST.					
KEYWORDS	Phytophthora sojae					
SOURCE	Phytophthora sojae					
ORGANISM	Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythaceae; Phytophthora.					
REFERENCE	1 (bases 1 to 615)					
AUTHORS	Tyler, B.					
TITLE	Tyler, B. Not Published					
JOURNAL	Unpublished (2003)					
COMMENT	Contact: Tyler B Tyler lab VBI 1880 Pratt Dr., Blacksburg, VA 24061, USA Tel: 540-231-7318 Email: bmttyler@vt.edu PCR Primers FORWARD: BK reverse primer BACKWARD: BK reverse primer Plate: 016 row: O column: 04					

Seq primer: BK reverse primer
 High quality sequence stop: 615.
 Location/Qualifiers
 1..615
 /organism="Phytophthora sojae"
 /mol_type="mRNA"
 /db_xref="taxon:67593"
 /clone="sMA016004"
 /tissue_type="mycelium"
 /cell_line="P6437"
 /dev_stage="mycelium"
 /lab_host="synthetic medium"
 /clone_lib="USDA-IFAPS:Expression of Phytophthora sojae genes during infection and propagation sMA"
 /notes="Vector: pBK-CMV; Site_1: EcoRI; Site_2: Xho1"

ORIGIN
 Query Match 21.6%; Score 71.4; DB 14; Length 615;
 Best Local Similarity 54.0%; Pred. No. 2.8e-05;
 Matches 169; Conservative 0; Mismatches 141; Indels 3; Gaps 1;
 QY 8 CAAAGTCCCGGTGATCTTCGGCGCAGCGCGGCTTCGAGCATCTGGTCGAACCGAAG 67
 Db |||||
 QY 69 CGAGGTGACCTCAAGTTCAGCTCGCGACACTCCATCAAGGAGTGGAGGCCAAGA 128
 Db |||||
 QY 68 CGGAGTATCGCTCATGGAAGCGCGCTTCTGAACCGCGGTGGACGGTATCGAAGCGTTT 127
 Db |||||
 QY 129 CGGCGATGTCCATCTCGGACGTGGCCACGCCAACGACATCGAC---CTCGAAGGGCGT 185
 QY 128 GCGGGGCGCGCTGTGCTCGCGCCACGTCGACGTTTACGTTGSCCGGAGTGGCTAGATG 187
 Db |||||
 QY 186 GCAGTGTCTCATGGCTGTCTCACGTGCACGTGATCTGGAGGACCCGTTGTCGACG 245
 QY 188 CGTGAAACCGCGAGTAGACCGAAGACGAATGTCGATGTCGTAGCGGAACGTCGCG 247
 Db |||||
 QY 246 AGCTGAGGAGGCGCTCGAGGACGAGGAGACATGCTGGACATGCGCTTCGCGCTCACGC 305
 QY 248 CGCATTCGCGGCTGCTGTCAGATCCGCTTACGACCTGTCGACGCGCTGACCTGG 307
 Db |||||
 QY 306 ACACGTCGATAGGCTGCGAGTGTCTGTGACGAGGCTTCGAGGCGACGCGTGA 365
 QY 308 AATGCGGAAGGC 320
 Db |||||
 QY 366 CGTGCCCAAGGC 378

RESULT 2
 AQ651128 578 bp DNA linear GSS 22-JUN-1999
 LOCUS Sheared DNA-20N12.TF Sheared DNA Trypanosoma brucei genomic clone
 DEFINITION Sheared DNA-20N12, genomic survey sequence.
 ACCESSION AQ651128
 VERSION AQ651128.1 GI:5144314
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei
 ORGANISM Trypanosoma brucei
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.
 1 (bases 1 to 578)
 El-Sayed,N., Zhao,S., Zhao,H., Gili,S., Suh,E., Malek,J., Fujii,C., Gerrard,C., Leach,V., de Jong,P., Ullis,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.
 Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library
 Unpublished (1999)
 Other GSSs: Sheared DNA-20N12.TF
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ARCC. Sheared DNA end sequences search page: <http://www.tigr.org/tdb/mdb/tbdb/>.
 Seq primer: M13-Forward
 Class: shotgun.
 Location/Qualifiers
 1..578
 /organism="Trypanosoma brucei"
 /mol_type="genomic DNA"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-20N12"
 /clone_lib="Sheared DNA-20N12"
 /note="Vector: pUC18; Site 1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. in Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

ORIGIN
 Query Match 21.3%; Score 70.2; DB 28; Length 578;
 Best Local Similarity 51.6%; Pred. No. 5e-05;
 Matches 159; Conservative 0; Mismatches 149; Indels 0; Gaps 0;
 QY 14 TCCGCGTATCTTCGCGCAGCGCGGCTTCGAGCATCTGGTCGAACCGAAGCGGAG 73
 Db |||||
 QY 193 TCCGCGTCAATGTGACACCGCTGAGGGGAAAAGATTACATTCTCTGCCCAAGTGGTT 252
 QY 74 TATCGCTCATGGAAGCGCGCTTCTGAACGCGTGGACGTTATCGAAGCGTTTTCGCGG 133
 Db |||||
 QY 253 TAAATTGATGGAGCGCTTCGTGACGTTGCTCGGTGGACATTGAAGCCGCTTGCACG 312
 QY 134 GCGGCTGTGCTGCGCACGTCGTCACGTTTACGTTGCCCCCGAGTGGCTAGATGCGGTGA 193
 Db |||||
 QY 313 GGACATGCGCTGTAGCATTGTCACTTCCTTCGGAAGAGGACTTTGGGAAACTGA 372
 QY 194 AACGCGGAGTGAGACGGAAGCAATGTCGATTGCGTAGCGGAACGTCGCGCGCAT 253
 Db |||||
 QY 373 CAGCGGATCGAGATGAATGACATGTGGACCTCGCGCCGANGTGAATCCGACGT 432
 QY 254 CGCGGCTGTCTGCCAGATCGGCTTACCGACCTGCTCGACGCGCTGACCTGGAATGC 313
 Db |||||
 QY 433 CGCGGCTCGGCTGCCAGTAAAGCTCAGCAAGAACTTGATGGATTACGCTGCAGATGC 492
 QY 314 CGAAGGCA 321
 Db |||||
 QY 493 CGAGTGA 500

RESULT 3
 CF241493/c
 LOCUS CF241493
 DEFINITION AGENCOURT 15113448 NICHG_XGC_Emb6 Silurana tropicalis cDNA clone IMAGE:6952976 3', mRNA sequence.
 ACCESSION CF241493
 VERSION CF241493.1 GI:33444701
 KEYWORDS EST.
 SOURCE Silurana tropicalis (western clawed frog)
 ORGANISM Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Silurana.
 1 (bases 1 to 917)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL

COMMENT

Contact: Daniela S. Gethard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Robert M. Grainger
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone Sequencing by: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14672 row: k column: 23
High quality sequence start: 8
High quality sequence stop: 752.
Location/Qualifiers
1. .917
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:6992976"
/tissue_type="neurula"
/dev_stages="embryo, stages 14-19"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHX XGC Emb6"
/note="Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.1 kb. Constructed by Invitrogen. Note: this is a
Xenopus Gene Collection (XGC) library."

ORIGIN

Query Match 20.1%; Score 66.4; DB 14; Length 917;
Best Local Similarity 60.2%; Pred. No. 0.00037;
Matches 109; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 116 TCGAACCGTTTCGGGGGGCGCTGCTGGCCGACGCGAAGATGCTCGATTGCGTAG 175
DB 800 TGAAGGCGCCGTGTAATCTCTGCTGCTCCACGTCGACGTTTACGTGAATCTG 741
QY 176 AGTGCTAGATGCTGTAACCGCCGAGTGAGACCGAAGATGCTCGATTGCGTAG 235
DB 740 AGTTTTCGATAAATCTCCAGAGCCGATGAAGGAGGATGATGCTGACATGSCAC 681
QY 236 CGGAACGTGCGCGATTCGGCGCTGCTGCGCAGATCCGCTTACGACCTGTCGACG 295
DB 680 GTTTCCTTCAGAGAACTCCGCGCTGCGCTGCGCAGATCATCTCACCAGAACTGAACG 621
QY 296 G 296
DB 620 G 620

RESULT 4

AL859568
LOCUS
DEFINITION AL859568 XGC-egg Silurana tropicalis cDNA clone TEGG060p07 5', mRNA
sequence.
ACCESSION AL859568
VERSION AL859568.2 GI:38655003
KEYWORDS EST.
SOURCE Silurana tropicalis (western clawed frog)
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 624)
REFERENCE
Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL Unpublished (2003)
COMMENT On Sep 15, 2002 this sequence version replaced gi:22979749.
Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001

TROPICALIS_SEQUENCE_ID: TEGG060p07.plkSP6
Sequencing Primer: SP6

This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.

cDNA was oligo dt primed from sug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue.

FEATURES
source

Location/Qualifiers
1. .624
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TEGG060p07"
/dev_stage="egg"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-egg"
/notes="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"

ORIGIN

Query Match 19.9%; Score 65.8; DB 9; Length 624;
Best Local Similarity 60.2%; Pred. No. 0.00045;
Matches 109; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 116 TCGAACCGTTTCGGGGGGCGCTGCTGGCCGACGCGAAGATGCTCGATTGCGTAG 175
DB 260 TGAAGGCGCCGTGTAATCTCTGCTGCTCCACGTCGACGTTTACGTGAATCTG 319
QY 176 AGTGCTAGATGCTGTAACCGCCGAGTGAGACCGAAGATGCTCGATTGCGTAG 235
DB 320 AGTTTTCGATAAATCTCCAGAGCCGATGAAGGAGGATGATGCTGACATGSCAC 379
QY 236 CGGAACGTGCGCGATTCGGCGCTGCTGCGCAGATCCGCTTACGACCTGTCGACG 295
DB 380 CGTTGCTTCAGGAACTCCGCGCTGCGCTGCGCAGATCATCTCACCAGAACTGAACG 439
QY 296 G 296
DB 440 G 440

RESULT 5

EX733466
LOCUS
DEFINITION EX733466 XGC-tadpole Silurana tropicalis cDNA clone TTPA053b02 5',
mRNA sequence.
ACCESSION EX733466
VERSION EX733466.1 GI:38406207
KEYWORDS EST.
SOURCE Silurana tropicalis (western clawed frog)
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 714)
REFERENCE
Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL Unpublished (2003)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TTPA053b02.plkSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.

CDNA was oligo dT primed from Sug of poly A+ RNA from tadpole embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.
 Vector: pCS107; Site 1: EcoRI; Site 2: NotI
 Host: Escherichia coli DH10B.
 Location/Qualifiers

FEATURES

source

1. 714
 /organism="Silurana tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="Tpa053b02"
 /dev_stage="tadpole (stage 35-40)"
 /lab_host="E. coli DH10B"
 /clone_lib="XGC-tadpole"
 /notes="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dT primed from Sug of poly A+ RNA from tadpole embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end"

ORIGIN

Query Match 19.9%; Score 65.8; DB 13; Length 714;
 Best Local Similarity 60.2%; Pred. No. 0.00047;
 Matches 109; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 QY 116 TCGAAGCCGTTTCGGGGGGCCCTGCTGCTGCCACGCGCCACGTTTACGTTGGCCCG 175
 Db 439 TGAAGGCGCTGTGAATCTCCTGCGCTCTCCACGTCGACGTTTACGTGAATACG 498
 QY 176 AGTGGCTAGATCGCTGAACCCCGAGTGCAGACCGAAGCAAAATGCTCGATTGGTAG 235
 Db 499 AGTTTTTCGATAAATCTTCAGAGCCCGATGAAGGAGGATGATGCTGGACATGGCAC 558
 QY 236 CGGAAGCGCGCGCATTCGCGCTGTCCTGCCAGATCCGCTTACGACCTGCTGACG 295
 Db 559 CGTTGCTTCAGGAGAACTCCCGCTCGGCTGCAGATCATTCTCACGAAGAAGTGAACG 618
 QY 296 G 296
 Db 619 G 619

RESULT 6
 CF343736
 LOCUS
 DEFINITION AGENCOURT 15225932 NICHHD_XGC_Swb1 Silurana tropicalis cDNA clone
 IMAGE:6999635 5', mRNA sequence.
 CF343736
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Silurana tropicalis (western clawed frog)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Silurana.

REFERENCE
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics / NIH
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-@email.nih.gov
 Tissue Procurement: Rob Granger, University of Virginia
 cDNA Library Preparation: Open Biosystems
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM14690 row: a column: 10
 High quality sequence stop: 681.
 Location/Qualifiers

FEATURES

Location/Qualifiers

source

1. 762
 /organism="Silurana tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="IMAGE:6999635"
 /tissue_type="whole body"
 /clone_lib="NICHHD_XGC_Swb1"
 /notes="Vector: pExpress-1; Site 1: EcoRV; Site 2: NotI; Bulk tissue was collected from a whole 10 month old male from the F6 strain. 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was cloned into the Not I and EcoRV sites of pExpress-1. Library was size-selected for >1.5 kb fragments for an average insert size of 1.92 kb. A normalized version of this library is also available (NICHHD_XGC_Swb1N). Library was constructed by Open Biosystems (Huntsville, AL)."

ORIGIN

Query Match 19.9%; Score 65.8; DB 14; Length 762;
 Best Local Similarity 60.2%; Pred. No. 0.00048;
 Matches 109; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 QY 116 TCGAAGCCGTTTCGGGGGGCCCTGCTGCTGCCACGCGCCACGTTTACGTTGGCCCG 175
 Db 500 TGAAGGCGCTGTGAATCTCCTGCGCTCTCCACGTCGACGTTTACGTGAATACG 559
 QY 176 AGTGGCTAGATCGCTGAACCCCGAGTGCAGACCGAAGCAAAATGCTCGATTGGTAG 235
 Db 560 AGTTTTTCGATAAATCTTCAGAGCCCGATGAAGGAGGATGATGCTGGACATGGCAC 619
 QY 236 CGGAAGCGCGCGCATTCGCGCTGTCCTGCCAGATCCGCTTACGACCTGCTGACG 295
 Db 620 CGTTGCTTCAGGAGAACTCCCGCTCGGCTGCAGATCATTCTCACGAAGAAGTGAACG 679
 QY 296 G 296
 Db 680 G 680

RESULT 7

CF376289
 LOCUS
 DEFINITION AGENCOURT 15350961 NICHHD_XGC_Swb1N Silurana tropicalis cDNA clone
 IMAGE:7004712 5', mRNA sequence.
 CF376289
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Silurana tropicalis (western clawed frog)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Silurana.
 REFERENCE
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics / NIH
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-@email.nih.gov
 Tissue Procurement: Rob Granger, University of Virginia
 cDNA Library Preparation: Open Biosystems
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM14702 row: h column: 22
 High quality sequence stop: 661.
 Location/Qualifiers
 1. 880
 /organism="Silurana tropicalis"

/mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="IMAGE:7004712"
 /tissue_type="whole body"
 /clone_lib="NICHD XGC Swb1N"
 /note="Vector: pExpress-1; Site 1: EcoRV; Site 2: NotI;
 Bulk tissue was collected from a whole 10 month old male
 from the P6 strain. 1st strand cDNA was primed with a Not
 I - oligo(dT) primer. double-stranded cDNA was cloned into
 the NotI and EcoRV sites of pExpress-1. Library was
 size-selected for >1.5 Kb fragments for an average insert
 size of 1.92 Kb. Library was normalized to Cot5 with a
 180-fold reduction of actin. A non-normalized version of
 this library is also available (NICHD XGC Swb1). Library
 was constructed by Open Biosystems (Huntsville, AL)."

ORIGIN

Query Match 19.9%; Score 65.8; DB 14; Length 880;
 Best Local Similarity 60.2%; Pred. No. 0.0005;
 Matches 109; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 QY 116 TCGAAGCGCTTGGGGGGCGCTGTGCTCGCCAGCTGCGCAGCTTTACGTTGGCCCG 175
 Db 458 TCGAAGCGCGCTGTGAATCTCTCCGCTGCTCCAGCTGCCAGCTTAGTGATCTG 517
 QY 176 AGTGCTAGATGCGCTGAACCGCGGAGTGAGACCGAAGAGAAATGCTGATGCGTAG 235
 Db 518 AGTTTTCGATAAACTTCAGAGCCGATGAAGAGGAGGATGATGCTGGACATGGCAC 577
 QY 236 CGGAACGTGCGCGCATTCGCGGCTGCTCCAGATCCGCTTACCGACCTGCTCGACG 295
 Db 578 CGTTGCTCAGAGAACTCCCGGCTGCGCTGCCAGATCATCTCCCGAAGAACTGAACG 637
 QY 296 G 296
 Db 638 G 638

RESULT 8

CF241492
 LOCUS
 DEFINITION AGENCOURT 15113421 NICHD XGC Emb6 Silurana tropicalis cDNA clone
 IMAGE:6992976 5', mRNA sequence.
 CF241492
 CF241492.1 GI:33444700

ACCESSION

VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Silurana tropicalis (western clawed frog)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Silurana.

REFERENCE

1 (bases 1 to 913)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892

AUTHORS

Blag. 31 Rm10A07 Bethesda, MD 20892
 Email: gcapsb-remail.nih.gov
 Tissue Procurement: Robert M. Grainger
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Agencourt Bioscience Corporation
 Clone distribution by: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM4672 row: k column: 23
 High quality sequence stop: 712.

FEATURES

source
 Location/Qualifiers
 1..913
 /organism="Silurana tropicalis"
 /mol_type="mRNA"

/db_xref="taxon:8364"
 /clone="IMAGE:6992976"
 /tissue_type="neural"
 /dev_stage="embryo, stages 14-19"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NICHD XGC Emb6"
 /note="Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV;
 Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 2.1 kb. Constructed by Invitrogen. Note: This is a
 Xenopus Gene Collection (XGC) library."

ORIGIN

Query Match 19.9%; Score 65.8; DB 14; Length 913;
 Best Local Similarity 60.2%; Pred. No. 0.0005;
 Matches 109; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 QY 116 TCGAAGCGCTTGGGGGGCGCTGTGCTCGCCAGCTGCGCAGCTTTACGTTGGCCCG 175
 Db 458 TCGAAGCGCGCTGTGAATCTCTCCGCTGCTCCAGCTGCCAGCTTAGTGATCTG 517
 QY 176 AGTGCTAGATGCGCTGAACCGCGGAGTGAGACCGAAGAGAAATGCTGATGCGTAG 235
 Db 518 AGTTTTCGATAAACTTCAGAGCCGATGAAGAGGAGGATGATGCTGGACATGGCAC 577
 QY 236 CGGAACGTGCGCGCATTCGCGGCTGCTCCAGATCCGCTTACCGACCTGCTCGACG 295
 Db 578 CGTTGCTCAGAGAACTCCCGGCTGCGCTGCCAGATCATCTCCCGAAGAACTGAACG 637
 QY 296 G 296
 Db 638 G 638

RESULT 9

BH195739/c
 LOCUS
 DEFINITION TC3-33B9.TV TC3 Trypanosoma cruzi genomic clone TC3-33B9, genomic
 survey sequence.
 BH195739
 BH195739.1 GI:16363721

ACCESSION

VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Trypanosoma cruzi
 Trypanosoma cruzi
 Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.

REFERENCE

1 (bases 1 to 525)
 Kluge, S., Edwards, K.E., Nilsson, D., Bontempi, E.J., Myler, P.,
 Stuart, K., Ghedin, E., El-Sayed, N.M. and Andersson, B.
 Clustering and analysis of BAC-end and GSS sequences from
 Trypanosoma cruzi

JOURNAL

Unpublished (2001)
 Contact: Bjorn Andersson
 Department of Genetics and Pathology
 Uppsala University
 Rudbeck Laboratory, SE-751 85, Uppsala, Sweden
 Tel: 46 18 471 4107
 Fax: 46 18 471 4808
 Email: bjorn.andersson@genpat.uu.se
 Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
 TC3. For clone availability, please contact Dr. Bjorn Andersson at
 Uppsala University (bjorn.andersson@genpat.uu.se).
 Seq primer: T7
 Class: BAC ends.

FEATURES

source
 Location/Qualifiers
 1..525
 /organism="Trypanosoma cruzi"
 /mol_type="genomic DNA"
 /strain="CL Brener"
 /db_xref="taxon:5693"
 /clone="TC3-33B9"
 /clone_lib="TC3"
 /note="Vector: pBelobAC11; Site 1: Hin dIII; Constructed
 for Uppsala University by Marie-Christine Le Paslier in

the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBlueoAclI digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN

Query Match 19.9%; Score 65.6; DB 28; Length 525;
Best Local Similarity 54.0%; Pred. No. 0.00047; Indels 0; Gaps 0;
Matches 134; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 68 CGGAGTATCGTCAATGGAAGCGCGCTTCTGAAGCGGTGAGCGATCGAAGCGCTT 127
|||||
Db 508 CGGGCATTACCTCATGAGGCAATACGGGACTTGGGAAGACTTGACATTGAGCGCGAT 449
|||||

QY 128 GCGGGGCGCTGTGCTCGCCAGCTGCCAGCTTTACGTTGCGCCCGAGTGGCTAGATG 187
|||||
Db 448 GTATGGAATTGTGCTCGAGCACCCTGCCAGCTGATCTCCGAGGAGGACTTTGAGA 389
|||||

QY 188 CGGTAAACCGCGAGTGAACGAGAGAAATGCTCGATTGCGTAGCGGAACGTGGGC 247
|||||
Db 388 AGCTAGTGGCCCTCAGAAGAGGAGTGGACATGCTGACCTGGCGCATCCGTCACAA 329
|||||

QY 248 CGCATTCGGGCTGTCTCCAGATCGGCTTACCGACCTGCTCGAGCGCCTGACCTGG 307
|||||
Db 328 AGACATCAGCATATCGTCCAAATTCAGCTTACGAGCGCGCTCGATGATTAACCGTGA 269
|||||

QY 308 AACTGCGG 315
|||||
Db 268 AGTTGCGG 261
|||||

RESULT 10

BM643564 667 bp mRNA linear EST 26-FEB-2002
LOCUS 17000687312521 A.Gam.ad.cdNA1 Anopheles gambiae cdNA clone
DEFINITION 19600449629190 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

ORIGIN

Query Match 19.9%; Score 65.6; DB 12; Length 667;
Best Local Similarity 56.5%; Pred. No. 0.00051;
Matches 122; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 105 CGTGACGGTATCGAAGCCGTTTGGGGGGCGCTGTGCTGCGCCACGTCGACGTTTA 164
|||||
Db 337 CGGTGTCGATGAGAGGGCGCTTGGAGGCGCTCGCTGGCTGACCACTGTCACTGTA 396
|||||

QY 165 CGTTGGCCCGAGTGGCTAGATCGGTGAAACCGCCGAGTGAGACGGAACGAAATGCT 224
|||||
Db 397 CGTGACGACGATATCTCGATCGTCTGGCGAGCCGGAGGAGGAGGACGATCTGCT 456
|||||

QY 225 CGATTGCTAGCGAAGTGGCGGATTCGGGGCTGTCTGCGAGATCGCGCTTACCGA 284
|||||
Db 457 CGATATGSCACCGCTTCTGCGCGAAACTCCCGCTCGGCTGCCAGATAGTGTGAGAA 516
|||||

QY 285 CTTGCTGACGCGCTGACCTGGAACCTGCCGAAGGC 320
|||||
Db 517 GGACCTGGAAGGATGCGGTTCAGCTTCGCGAGGC 552
|||||

RESULT 11

CF220100 979 bp mRNA linear EST 04-AUG-2003
LOCUS 150403112 NICHHD XGC Emb5 Silurana tropicalis cdNA clone
DEFINITION IMAGE:6988704 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

CF220100 979 bp mRNA linear EST 04-AUG-2003
AGENCOURT 150403112 NICHHD XGC Emb5 Silurana tropicalis cdNA clone
IMAGE:6988704 5', mRNA sequence.

CF220100
CF220100.1 GI:33420808
EST.

Silurana tropicalis (western clawed frog)
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Silurana.

1 (bases 1 to 979)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics / NIH
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: gcrabbs@mail.nih.gov
Tissue Procurement: Robert M. Grainger
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM14661 row: 1 column: 23
High quality sequence stop: 687.
Location/Qualifiers
1..979
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:6988704"
/tissue_type="gastrula"
/dev_stage="embryo, stages 10-13"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHHD XGC Emb5"
/note="Vector: PCMV-SF0RT5.1; Site 1: NotI; Site 2: EcoRV;
Cloned unidirectionally. Primer: Cligo dt. Average insert
size 2.0 kb. Constructed by Invitrogen. Note: This is a
Xenopus Gene Collection (XGC) library."

FEATURES

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

BM643564 667 bp mRNA linear EST 26-FEB-2002
LOCUS 17000687312521 A.Gam.ad.cdNA1 Anopheles gambiae cdNA clone
DEFINITION 19600449629190 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

source

source

source

source

BM643564 667 bp mRNA linear EST 26-FEB-2002
LOCUS 17000687312521 A.Gam.ad.cdNA1 Anopheles gambiae cdNA clone
DEFINITION 19600449629190 5', mRNA sequence.

BM643564
BM643564.1 GI:18943075
EST.

Anopheles gambiae (African malaria mosquito)
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
1 (bases 1 to 667)
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004ABS row: C column: 12
Seq primer: M13 Reverse.
Location/Qualifiers
1..667
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449629190"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="A.Gam.ad.cdNA1"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole

ORIGIN

Query Match 19.6%; Score 64.8; DB 14; Length 979;
 Best Local Similarity 59.7%; Pred. No. 0.00084;
 Matches 108; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 116 TCGAAGCGTTTGGGGGGCCCTGTGCTCGCCACGTCGACGTTTACGTTGGCCCCG 175
 DB 555 TGAAGGGCGCTGTGAATCTCCCTGGGCTCTCCACGTGCCACGTTTACGTGAATACTG 614

QY 176 AGTGGCTAGATGCGCTGAACCGCGGAGTGAGACCGAGACGAAATGCTCGATTGCGTAG 235
 DB 615 AGTTTTCGATAAACTTCAGAGGCCGATGAAGAGGAGGATGATATGCTGACATGGCAC 674

QY 236 CGGAACGTGCGCGCCGCAATTCGGGGGTGTCTCTGCCAGATCCGGCTTACCGACCTGTCGACG 295
 DB 675 CGTTGCTTCANGAAGAACTCCCGGCTCGGCTGCCAGATCAATTCACCGAGAGCTGAACG 734

QY 296 G 296
 DB 735 G 735

RESULT 12

CB405072/c
 LOCUS CB405072 483 bp mRNA linear EST 15-MAY-2003
 DEFINITION OSTRO36H11_1 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
 ACCESSION CB405072
 VERSION CB405072.1 GI:30746799
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans

REFERENCE
 AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 483)
 Reboul,J., Vaglio,P., Rual,J.P., Lamesch,P., Martinez,M.,
 Armstrong,C.M., Li,S., Jacotot,L., Bertin,N., Janky,R., Moore,T.,
 Hudson,J.R., Hartley,J.L., Brasch,M.A., Vandenhaute,J., Boulton,S.,
 Endress,G.A., Jenna,S., Chevet,E., Papasotiropoulos,V.,
 Tolias,P.P., Ptacek,J., Snyder,M., Huang,R., Chance,M.R., Lee,H.,
 Doucette-Stamm,L., Hill,D.E. and Vidal,M.

TITLE C. elegans ORFeome version 1.1: experimental verification of the
 genome annotation and resource for proteome-scale protein
 expression

JOURNAL

COMMENT Nat. Genet. (2003) In press

Contact: Vidal M
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739

Email: Marc.Vidal@fci.harvard.edu
 Sequence tag of Gateway entry clones. The primers used were
 designed on the predicted protein encoding ORF. C. elegans ORFeome
 cloning project : Contact david_hill@fci.harvard.edu or
 marc_vidal@fci.harvard.edu
 POLYA=No.

FEATURES

source

Location/Qualifiers

1. 483
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_lib="AD-wrmcDNA"
 /note="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"

ORIGIN

Query Match 19.5%; Score 64.4; DB 14; Length 483;
 Best Local Similarity 57.4%; Pred. No. 0.00084;
 Matches 116; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 115 ATGAAGCGTTTGGGGGGCCCTGTGCTCGCCACGTCGACGTTTACGTTGGCCCC 174
 DB 226 ATGAGGGAGCTTGGAGGCATCACTTGCCTCTCGAGTGCACGCTACGTCGATCCA 167

QY 175 GAGTGGCTAGATGCGCTGAACCGCGGAGTGAGACCGAGACGAAATGCTCATTCGCTA 234
 DB 166 GCGTTCAGAACCAAGCTCCCGAGCGGCTAGAGAGGAGACGATATGCTGGATATGGCT 107

QY 235 GGGGAACGTGCGCGCCGCAATTCGGGGGTGTCTCTGCCAGATCCGGCTTACCGACCTGTCGAC 294
 DB 106 CCGGCGCTCAAGACAACTCACGGCTCGGCTCTCAATTGTGCTCACGAAAGAGCTCGAC 47

QY 295 GCGCTGACCTGGAACCTGCCGA 316
 DB 46 GGAATCACTGTAACCCCTGCCGA 25

RESULT 13

CB385967
 LOCUS CB385967 488 bp mRNA linear EST 15-MAY-2003
 DEFINITION OSTRO36H11_1 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
 ACCESSION CB385967
 VERSION CB385967.1 GI:30727677
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans

REFERENCE
 AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 488)
 Reboul,J., Vaglio,P., Rual,J.P., Lamesch,P., Martinez,M.,
 Armstrong,C.M., Li,S., Jacotot,L., Bertin,N., Janky,R., Moore,T.,
 Hudson,J.R., Hartley,J.L., Brasch,M.A., Vandenhaute,J., Boulton,S.,
 Endress,G.A., Jenna,S., Chevet,E., Papasotiropoulos,V.,
 Tolias,P.P., Ptacek,J., Snyder,M., Huang,R., Chance,M.R., Lee,H.,
 Doucette-Stamm,L., Hill,D.E. and Vidal,M.

TITLE C. elegans ORFeome version 1.1: experimental verification of the
 genome annotation and resource for proteome-scale protein
 expression

JOURNAL

COMMENT Nat. Genet. (2003) In press

Contact: Vidal M
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739

Email: Marc.Vidal@fci.harvard.edu
 Sequence tag of Gateway entry clones. The primers used were
 designed on the predicted protein encoding ORF. C. elegans ORFeome
 cloning project : Contact david_hill@fci.harvard.edu or
 marc_vidal@fci.harvard.edu
 POLYA=No.

FEATURES

source

Location/Qualifiers

1. 488
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_lib="AD-wrmcDNA"
 /note="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"

ORIGIN

```

Query Match      19.5%; Score 64.4; DB 14; Length 488;
Best Local Similarity 57.4%; Pred. No. 0.0084;
Matches 116; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 115 ATCGAGCCGTTTGGCGGCGCCTGTGCTGCGCCACGTCGACGTTTACGTTGGCCCC 174
DB 239 ATGAGGGAGCTTGCAGGAGCATCACTTGGCTGCTGACGTGCCAGCTTACGTCGATCA 298
QY 175 GAGTGTAGATCGCTGAAACCGCGAGTGTAGACCGAAGAGAAATGCTCGATTGCGTGA 234
DB 299 GCCTTCCAGAAACAAGCTCCGGAGCGCTAGAGAGAGAGAGATATGCTGATATGGCT 358
QY 235 GCGGACGTCGCGCGCATTCGGGCTGTCTGCGAGATCGGCTTACCGACCTGTCGAC 294
DB 359 CCGGCGCTCAAGGACAACTCACGCTCGGCTGTCAAATTTGTGCTACGAAAGAGCTCGAC 418
QY 295 GGCCTGACCTCGAACTGCGGA 316
DB 419 GGAATCACTGTACCTGCGGA 440

RESULT 14
BI626767
LOCUS
DEFINITION
  RH67819.5prime RH Drosophila melanogaster normalized Head pFlc-1
  Drosophila melanogaster cDNA clone RH67819 5 similar to fdxh;
  FBan0004205 GO: [ferredoxin (GO:0008042); electron transfer
  (GO:0005489)] located on: 3L 67B1-67B1;; 08/24/2001, mRNA sequence.
ACCESSION
  BI626767
VERSION
  BI626767.1 GI:15522292
KEYWORDS
  EST.
SOURCE
  Drosophila melanogaster (fruit fly)
ORGANISM
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 577)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Friese,E.,
  George,R., Gonzalez,R., Guoan,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Munoo,J., Pacleb,J., Paragas,V., Park,S.,
  Prounanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RH Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  hit genomic AB003552; arm:3L [9109492,9396275]
  estimated-cyto:67A1-67B5; 08/24/2001
  Plate: RH.678 row: B column: 7
  High quality sequence stop: 379.
FEATURES
  source
    1..577
    /organism="Drosophila melanogaster"
    /mol_type="mRNA"
    /db_xref="taxon:7227"
    /clone="RH67819"
    /sex="male and female"
    /dev_stage="Adult"
    /lab_host="DHS-alpha Tona"
    /clone_lib="RH Drosophila melanogaster normalized Head
    pFlc-1"
    /note="Organ: head; Vector: pFlc1; Site 1: XhoI; Site 2:
    BamHI; Library was kindly generated by Piero Carninci at
    the RIKEN. The library was normalized and excised using
    Cre recombinase. Plasmid cDNA library."
ORIGIN
  Query Match      19.0%; Score 62.8; DB 12; Length 577;

```

```

Best Local Similarity 56.2%; Pred. No. 0.0019;
Matches 118; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 111 CGGTATCGAAGCGTTTGGCGGCGCCTGTGCTGCGCCACGTCGACGTTTACGTTGG 170
DB 282 CGAGATCGAAGGCGCTGTGAGGCTTGGCTGCGCTGCACCACTGTACGTTGACGTCCA 341
QY 171 CCGGAGTGTGATGCGCTGAAACCGCGGAGTGTAGACCGAAGAGAAATGCTCGATTG 230
DB 342 GCATGATTACCTGTCAGAAAGTTAAAGAGGCGCGAGGAGGATGACGACCTGCTGGATAT 401
QY 231 CGTAGCGAAGCGTGCAGCGCATTCGGGCTGTCTGCGAGATCGGCTTACCGACCTGCT 290
DB 402 GGCCTATTTCTGCGCGAAGCTCCAGCTCGGCTGTGAGTACTCTCTCAGACAGATAT 461
QY 291 CGACGGCTGACCTGGAACCTGCGGAAGGC 320
DB 462 GGACGGCATGGAAGCTGGAGCTGCCAAGGC 491

RESULT 15
AI562398
LOCUS
DEFINITION
  TENS2389 T. cruzi epimastigote normalized cDNA library Trypanosoma
  cruzi cDNA clone 2389 5', mRNA sequence.
ACCESSION
  AI562398
VERSION
  AI562398.1 GI:4513743
KEYWORDS
  EST.
SOURCE
  Trypanosoma cruzi
  Trypanosoma cruzi
  Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
  Trypanosoma; Schizotrypanum.
REFERENCE
  1 (bases 1 to 632)
  Verdun,R.E., Di Paolo,N.C., Urmenyi,T.P., Rondinelli,E.,
  Frasch,A.C.C. and Sanchez,D.O.
  Gene discovery through expressed sequence tag sequencing in
  Trypanosoma cruzi
  Infect. Immun. 66 (11), 5393-5398 (1998)
  99003155
  9784549
  Contact: Sanchez D.O.
  Instituto de Investigaciones Biologicas (Univ. Nac. de Gral
  San Martin)
  Av. Gral Paz entre Albarillos y Constituyentes, INTI edificio 24
  cp(1650) San Martin, Prov. de BS AS, Argentina
  Tel: (54-1)752-9639 or (54-1)752-0021
  Fax: (54-1)752-0021 or (54-1)752-9639
  Email: dsanchez@inti.gov.ar
  Seq primer: T7.
FEATURES
  source
    Location/Qualifiers
    1..632
    /organism="Trypanosoma cruzi"
    /mol_type="mRNA"
    /strain="Cl-Brenner"
    /db_xref="taxon:5693"
    /clone="2389"
    /cell_type="epimastigote"
    /clone_lib="T. cruzi epimastigote normalized cDNA library"
    /note="cDNA library constructed with oligo dt primed
    epimastigote mRNA and cloned in pT7318d phagemid with
    modified polylinker (PHARMACIA)"
ORIGIN
  Query Match      18.7%; Score 61.8; DB 9; Length 632;
  Best Local Similarity 56.7%; Pred. No. 0.0033;
  Matches 114; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 115 ATCGAAGCGTTTGGCGGCGCCTGTGCTGCGCCACGTCGACGTTTACGTTGGCCCC 174
DB 50 ATTGAGGCGCATGTATGGAACCTGTGCTGCGACACCTGCCACGTGATCTCCGGAG 109
QY 175 GAGTGGCTAGATGCGCTGAACCCCGAGTCGACCGAAGAGAAATGCTCGATTGCGTA 234

```


Db	110	GAGGACTTTGAGAAGCTGAGTGAGCCCTCAGAGACGAGGTGGACATGCTCGACTTGGCG	169
Qy	235	GCGGACGTGGCGCGCATTCGGGCTGTCTGCCAGATCGCCTTACCGACTGCTCGAC	294
Db	170	CCATCCGTCACAAAGACATCAGATTATCGTGCCAAATTCAGCTTACGGACGGGCTCGAT	229
Qy	295	GGCCTGACCCCTGGAACTGCCG	315
Db	230	GGTATTACCGTGAAGTTGCCG	250

Search completed: June 19, 2004, 07:19:52
Job time : 971.906 secs

Db 1431 GGGTCAGCCTGATGACAGGCTGAGTCTCCAAATGGTATCTACGATATTGTGCGTGAATGT 1490
QY 130 GGGGGCGCTGTGCTCTGCCACGTCGCCACGTTTACGTTGGCCCGAGTGGCTAGATCGG 189
Db 1491 GCGCGCAGCGCCAGCTGTGCCACCTGCCATGCTATGTGAACGAAGCGTTACGGACAAG 1550
QY 190 CTGAACCCCGCAGTGAGACCGAAGCAAAATGCTCGATTGGT---AGCGAAGCTGG 246
Db 1551 GTGCCCGCCCAACAGAGGGGAAATCGGCATGCTGGAGTGGTCAACGCCGAATCGAAG 1610
QY 247 CGCATTTCGCGCTGTCTGCCAGATCCGCCCTTACCGACCTGCTCGACGGGCTGACCCCTG 306
Db 1611 CCGAACAGCAGGCTGTCTGTGCCAGATCATCATGACGCCGAGCTGGATGTCGTGTC 1670
QY 307 GAACGCCCAAGGACAGTCATGA 330
Db 1671 GATGTTCCCGATAGGCAATGGTAA 1694

RESULT 2
US-09-252-991A-6336
; Sequence 6336 Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6336
; LENGTH: 345
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6336

Query Match 16.4%; Score 54; DB 4; Length 345;
Best Local Similarity 58.3%; Pred. No. 1.4e-05;
Matches 134; Conservative 0; Mismatches 90; Indels 6; Gaps 2;
QY 56 TCGAACCGAAGCGGAGTATCGTCTATGAAAGCGCGCTTCTGAACGCGCGTGGACGTA 115
Db 62 TCGAGCGGAAGCCCGCGCAACCATCTGATGCGCGCTGGCAATGGCATCGA---GA 118
QY 116 TCGAAGCCGTTTTCGGGGCGCCTGTGCTGCGCCACGTCGCCACGTTTACGTTGGCCCG 175
Db 119 TCGAGCATGCTGTGAGAAATCTTGGCTGCACTGCGCTGCGCAATGCTGTT 175
QY 176 ATGGCTAGATCGCTGAACCGCGAGTGAGACCGAAGCAAAATGCTCGATTGCTGAG 235
Db 176 AAGGCTCAGTCCATGAGCGGCTCGACAACTCGAAGACGACATGCTGGCAAGGCT 235
QY 236 CGGAAGCTGCGCGGATTCGCGCTGTCTGCGCAGATCGGCTTACCGAC 285
Db 236 GGGTCTGAGCCGATTCGCGCTATCTGCGCAGTATCTGCGAGCGATGTTGGCCCGAC 285

RESULT 3
US-09-252-991A-6047/c
; Sequence 6047 Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6047
; LENGTH: 756
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6047

Query Match 16.4%; Score 54; DB 4; Length 756;
Best Local Similarity 58.3%; Pred. No. 1.6e-05;
Matches 134; Conservative 0; Mismatches 90; Indels 6; Gaps 2;
QY 56 TCGAAACCGAAGCGGAGTATCGCTCATGAAAGCGCGCTTCTGAACGCGGTGGACGTA 115
Db 388 TCGAGCGAAGCCCGCGCAACCATCTGATGCGCGCTTGGCAATGGCATCGA---GA 332
QY 116 TCGAAGCCGTTTTCGGGGCGCCTGTGCTGCGCCACGTCGCCACGTTTACGTTGGCCCG 175
Db 331 TCGAGCATGCTGTGAGAAATCTTGGCTGCACTGCGACCTGCCACGTAATCGT---CCGCG 275
QY 176 AGTGGCTAGATGCGCTGAAACCGCGAGTGAGACCGAAGCAAAATGCTCGATTGCTAG 235
Db 274 AAGGCTCGACTCCATGGAGCGCTCGGAGCAACTCGAAGACGACATGCTGGCAAGGCT 215
QY 236 CGGAAGCTGCGCGCATTCGCGCTGTCTGCGCAGATCGGCTTACCGAC 285
Db 214 GGGTCTGAGAGCCCGATTCGCGCTATCTTCCAGGCACTGTTGGCCCGAC 165

RESULT 4
US-09-489-039A-4374
; Sequence 4374 Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 4374
; LENGTH: 363
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4374

Query Match 14.6%; Score 48.2; DB 4; Length 363;
Best Local Similarity 55.6%; Pred. No. 0.00048;
Matches 114; Conservative 0; Mismatches 88; Indels 3; Gaps 1;
QY 81 CATGGAAGCGCGCTTCTGAACGCGCTGGACGCTATCGAAGCGGTTTGGGGGGCGCGCTG 140
Db 105 CATTCGTGATGCGCGCTGCTGAGCGGTATCGAAATTAACACGCTCGGAAAAATCCCTG 164
QY 141 TGCTGCGCCACGTCGCACGTTTACGTTGGCCCGAGTGGCTAGATGCGCTGAAACCGCC 200
Db 165 TGCTGCAACCACTGCGCAC---TGCATCGTGCCTGAAGGTTTGACTCGCTGGCGGAAAG 221
QY 201 GAGTGAGACCGAAGCAATGCTGATGCGTAGCGGAACGTCGCGCGCATTCGGGCT 260
Db 222 TAGCGAAGATGAAGACGACATGCTGGAATAAGCTGGGGCTGGAACCTGACAGCGCTCT 281
QY 261 GTCTGCGCATCCGCTTACCGAC 285
Db 282 GAGTGGCAGCGCGGTCAACCGAC 306

RESULT 5
US-09-252-991A-6212/C
; Sequence 6212, Application US/09252991A
; Patent No. 8551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6212
; LENGTH: 2112
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6212

Query Match 13.0%; Score 43; DB 4; Length 2112;
Best Local Similarity 57.6%; Pred. No. 0.015;
Matches 117; Conservative 0; Mismatches 80; Indels 6; Gaps 2;

QY 56 TCGAACCAGGCGGAGTATCGCTCATGGAAGCGCGGTTCTGAACGGCGTGGACGGTA 115
DB 197 TCGAGGCGAAGCGCGGCGAACCATACTGATGCGCGTTGCGCAATGGCATCGA---GA 141
QY 116 TCGAAGCGGTTTGGGGGGCGCTGTGCTGCGCCAGTGCACGTTTACGTTGGCCCG 175
DB 140 TCGAGCATGCTGTGAAATCTTGGCCCTGCGACCTGCCACGTAATCGT---CCGCG 84
QY 176 AGTGGCTAGATGCGCTGAACCGCGAGTGAACCGAAGACGAAATGCTCGATTGCGTAG 235
DB 83 AAGCGCTCGACTCATCGAGCGCGTCCGACGAACTCGAAGACGACATGCTGGACAAGCGCT 24
QY 236 CGGAACGTGCGCGCATTCGCG 258
DB 23 GGGGCTGGAGCGCGATTCCGG 1

RESULT 6
US-09-252-991A-6417
; Sequence 6417, Application US/09252991A
; Patent No. 8551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6417
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6417

Query Match 12.8%; Score 42.2; DB 4; Length 405;
Best Local Similarity 59.7%; Pred. No. 0.019;
Matches 71; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 167 TTGCCCCCGAGTGGCTAGATCGCTGAACCGCGAGTGAGACCGAAGACGAATGCTCG 226
DB 1 TCGTCCGCGAAGGCGCTCGACTCCATGAGCGCTCCGACGAACTCGAAGACGACATGCTCG 60

QY 227 ATTGCTAGCGAAGCGTGGCGGCTTCCGGGCTGTCTCTGCCAGATCGCGCTTACCGAC 285
DB 61 ACAAGGCTGSGGTCTGGAGCCCGATTCCGGGCTATCTGCCAGGACGTCGTGGCGGAC 119

RESULT 7
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhnstock, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-227C-7

Query Match 12.8%; Score 42.2; DB 2; Length 44377;
Best Local Similarity 47.5%; Pred. No. 0.042;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 3 GACCCGAAGTTCGCGGTGATCTTCCGGCAGCGCGGCTTCGAGCATCTGTCGAAC 62
DB 29022 GCCCGCAGGTGCGCGTTCGCGGCTAGCGCCCGCGGCGTCACTTCGCGAGCGCCTCAT 29081
QY 63 CGAAGCGGAGTATCGCTCATGGAAGCGCGCGCTTTGAAACGCGTGGACGTTATCGAAGC 122
DB 29082 CGCCCTCGCATGTATACCGCGGCGAGATGGAAACGGAGGCGCGGCGCATCGTGT 29141

QY 123 CGTTTGGGGGGCGCTGTGCTGGCCAGCGTGCACGTTTACGTTGTCGCCCGCGAGTGCT 182
Db 29142 CGAGTGGCGCCCGCGTACCGGTGTCGCCGTGGCGACCGCGGTGCTCGCCCTGTGGGA 29201
QY 183 AGATCGGTGAACCGCGAGTGAACCGAAGACGAAATGCTCGATTGCTGAGCGGAACG 242
Db 29202 CGCGGCGCTGGCGCGCTGTGGTGGCGACACCGCGTGTCTGCCCGCGTCCCGGAGG 29261
QY 243 TCGCGCGGATTGCGCGGTGTCT 265
Db 29262 CTGGTCTACGCCCGCGCGCT 29284

RESULT 8
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Bargett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostek, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4437 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830

; Query Match 12.8%; Score 42.2; DB 2; Length 44377;

Best Local Similarity 47.5%; Pred. No. 0.042;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
QY 3 GACCGAAAGTCCGGTGAATCTTCGCCACGCGCGGCTTCGAGCATCTCTGGTGAAGC 62
Db 29022 GCCCGCCAGGTGCGGTGCGGTACGCGCCGCGGCGCTCAACTTCGCGGACGCCCTCAT 29081
QY 63 CGAAGCGGGAGTATCGCTCATGGAAGCGCGCTCTGAACGCGGTGAGACGGTATCGAAGC 122
Db 29082 CGCCCTCGGCATGTACCCGCGGAGCGGAGATGGGACGAGGGCGCGGACCGTCTGT 29141
QY 123 CGTTTCGGGGCGCTGTGCTTCGCCACGTCGCCACGTTTACGTTTGGCCCCGAGTGGCT 182
Db 29142 CGAGGTCCGCCCGCGGTCAACCGGTGTCGCCGACCGCGTCTCGGCTGTGGGA 29201
QY 183 AGATCGCTGAACCGCGAGTGAGACCGAAGACGAAATGCTCGATTGCTAGCGGAACG 242
Db 29202 CGCGGCGCTGGCGCCCGCTGTGCGTGGCCGACACCGGCTGTCCGCCCGTCCCGACG 29261
QY 243 TCGCGCGCATTCGCGGTGTCT 265
Db 29262 CTGGTCTACGCCCGCGCGCT 29284

RESULT 9
US-09-427-700-8
; Sequence 8, Application US/09427700
; Patent No. 6372489
; GENERAL INFORMATION:
; APPLICANT: AntiCancer
; APPLICANT: Zhao Ming
; TITLE OF INVENTION: METHOD AND MODEL FOR HAIR PIGMENTATION
; FILE REFERENCE: 31276-20017.00
; CURRENT APPLICATION NUMBER: US/09/427,700
; CURRENT FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: US 60/105,725
; PRIOR FILING DATE: 1998-10-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (16)...(450)
; OTHER INFORMATION: ORF 438 gene
; US-09-427-700-8

Query Match 11.9%; Score 39.2; DB 4; Length 459;
Best Local Similarity 50.0%; Pred. No. 0.12;
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 4 ACCGAAAGTCCGGTGAATCTTCGCCGAGCGCGGCTTCGAGCATCTGGTGAAGC 63
Db 121 ACCCCGAGGTCCCGGGAACCGCGCGGTTCGGCGCCCGCCCTTCGACGAGATC 180
QY 64 GAAGCGGAGTATCGCTCATGGAAGCGGCGCTTCTGAACGGGTGAGACGGTATCGAAGC 123
Db 181 TACAGGGCGCGCGATACAGGGCCCGGACCGGCGGCGGCCACACGCGCGC 240
QY 124 GTTTGGGGGCGCTGTGCTTCGCCACGTCGCCACGTTTACGTTGGCCCCGAGTGGCTA 183
Db 241 GGTACGCGGTGACCGTCAACCGCGCGGCGGCTACGCGCGGCGGCTTACGCGGTTCGTG 300
QY 184 GATGCGCTGAACCGC 199
Db 301 GACGGGTGCACTGC 316

RESULT 10
US-07-989-363-1
; Sequence 1, Application US/07989363

Patent No. 5385841
GENERAL INFORMATION:
APPLICANT: Ortega, Jose Daza
APPLICANT: Gill, Jose Antonio
APPLICANT: Garcia, Tomas Vigil
APPLICANT: Martin, Juan Francisco
TITLE OF INVENTION: PO438, A NEW CALCIUM-REGULATED PROMOTER
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Suite 701-E Columbia Square
STREET: 555 13th Street, N. W.
CITY: Washington
STATE: D. C.
COUNTRY: U. S.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/989,363
FILING DATE: 19921211
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Figg, E. Anthony
REGISTRATION NUMBER: 27,195
REFERENCE/DOCKET NUMBER: 1768-112A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 572 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 94..531
US-07-989-363-1

Query Match 11.9%; Score 39.2; DB 1; Length 572;
Best Local Similarity 50.0%; Pred. No. 0.12;
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
Qy 4 ACCGCAAGGTCGCGTGATCTTCGCGCAGCGCGCGCTTCGAGCATCTGGTCGAAACC 63
Db 202 ACCCGCGGTCCTCCCGGGAACCGCGCGCTTCGCGCGCGCGCTTCGAGCATC 261
Qy 64 GAAGCGGAGTATCGCTCATGGAAGCGCGCGCTTCGAAAGCGGCGTGGACGGTATCGAAGCC 123
Db 262 TACAAGGCGCGCGCGGATACAGGGCGCGGACGTCACCGACGCGCGCGCGCGCGC 321
Qy 124 GTTTGCGGGGCGCGCTGTGCTCGCGCAGCGTGCACCGTTTACGTTGCGCGCGCGCGCTA 183
Db 322 GGTACGCGCGTACGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 381
Qy 184 GATGCGTGAACCGC 199
Db 382 GACGCGTCGAACTGC 397

RESULT 11
US-08-264-526-1
Sequence 1, Application US/08264526
Patent No. 566099
GENERAL INFORMATION:
APPLICANT: Ortega, Antonio Daza
APPLICANT: Gill, Jose Antonio
APPLICANT: Garcia, Tomas Vigil

APPLICANT: Martin, Juan Francisco
TITLE OF INVENTION: PO438, A NEW CALCIUM-REGULATED PROMOTER
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Suite 701-E Columbia Square
STREET: 555 13th Street, N. W.
CITY: Washington
STATE: D. C.
COUNTRY: U. S.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/264,526
FILING DATE: 23-JAN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/989,363
FILING DATE: 11-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Figg, E. Anthony
REGISTRATION NUMBER: 27,195
REFERENCE/DOCKET NUMBER: 1768-119A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 572 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 94..531
US-08-264-526-1

Query Match 11.9%; Score 39.2; DB 1; Length 572;
Best Local Similarity 50.0%; Pred. No. 0.12;
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
Qy 4 ACCGCAAGGTCGCGTGATCTTCGCGCAGCGCGCGCTTCGAGCATCTGGTCGAAACC 63
Db 202 ACCCGCGGTCCTCCCGGGAACCGCGCGCTTCGCGCGCGCGCTTCGAGCATC 261
Qy 64 GAAGCGGAGTATCGCTCATGGAAGCGCGCGCTTCGAAAGCGGCGTGGACGGTATCGAAGCC 123
Db 262 TACAAGGCGCGCGCGGATACAGGGCGCGGACGTCACCGACGCGCGCGCGCGCGC 321
Qy 124 GTTTGCGGGGCGCGCTGTGCTCGCGCAGCGTGCACCGTTTACGTTGCGCGCGCGCGCTA 183
Db 322 GGTACGCGCGTACGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 381
Qy 184 GATGCGTGAACCGC 199
Db 382 GACGCGTCGAACTGC 397

RESULT 12
US-08-152-483B-8
Sequence 8, Application US/08152483B
Patent No. 5529909
GENERAL INFORMATION:
APPLICANT: Della-Cioppa, Guy
APPLICANT: Kumagai, Monto
TITLE OF INVENTION: TYROSINASE-ACTIVATOR
TITLE OF INVENTION: PROTEIN FUSION ENZYME
NUMBER OF SEQUENCES: 9

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pennie & Edmonds
;; STREET: 2730 Sand Hill Road
;; CITY: Menlo Park
;; STATE: California
;; COUNTRY: U.S.A.
;; ZIP: 94025
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent in Release #1.0,
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/152,483B
;; FILING DATE: No. 5529909ember 12, 1993
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 7/857,602
;; FILING DATE: March 30, 1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 923,692
;; FILING DATE: July 31, 1992
;; APPLICATION NUMBER: 600,244
;; FILING DATE: October 22, 1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 641,617
;; FILING DATE: January 16, 1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 737,899
;; FILING DATE: July 26, 1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Halluin, Albert P.
;; REGISTRATION NUMBER: 25,227
;; REFERENCE/DOCKET NUMBER: BIOG-20240/8129-040
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 854-3660
;; TELEFAX: (415) 854-3694
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1442
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA
;; DESCRIPTION: NO
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM:
;; IMMEDIATE SOURCE:
;; CLONE:
;; FEATURE:
US-08-152-483B-8

Query Match 11.9%; Score 39.2; DB 1; Length 1442;
Best Local Similarity 50.0%; Pred. No. 0.14;
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 4 ACCGCAAGTCGGGTGATCTTCGCGAGCGGGGTTTCGAGCATCTGTCGAAACC 63
Db 288 ACCCGGAGTCCCCGGGAACCGCGGTCGCGCGCCCGCCCTTCGACGAGATC 347
QY 64 GAAGCGGGAGTATCGCTCATGGAAGCGCGCGGTCTTCTGAACGGGGTGGACGGTATCGAAGCC 123
Db 348 TACAAGGCGCGCGGATACAGCGCGGACGGTCACCGACGGCGGGGTCACCGCGGC 407
QY 124 GTTTCGGGGGGCGCTGTGCTCCCTGCGCCACGTGCCACGTTTACGTGGCCCCGAGTGGCTA 183
Db 408 GGTACGCGCGGTGACGGGTACGCGCGCGCGCCATCACGCGCGCGGTTCGCGGTTCGTG 467

QY 184 GATCGCTGAAACGC 199
Db 468 GACGCGTCGAACTGC 483
RESULT 13
US-08-152-483B-6
; Sequence 6, Application US/08152483B
; Patent No. 5529909
; GENERAL INFORMATION:
; APPLICANT: Della-Gioppa, Guy
; APPLICANT: Kumagai, Monte
; TITLE OF INVENTION: TYROSINASE-ACTIVATOR
; TITLE OF INVENTION: PROTEIN FUSION ENZYME
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 2730 Sand Hill Road
; CITY: Menlo Park
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,483B
; FILING DATE: No. 5529909ember 12, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 7/857,602
; FILING DATE: March 30, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 923,692
; FILING DATE: July 31, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 600,244
; FILING DATE: October 22, 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 641,617
; FILING DATE: January 16, 1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 737,899
; FILING DATE: July 26, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: BIOG-20240/8129-040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4009
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE:
; DESCRIPTION: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM:
; IMMEDIATE SOURCE:
; CLONE:
; FEATURE:
US-08-152-483B-6

Query Match 11.9%; Score 39.2; DB 1; Length 4009;

Best Local Similarity 50.0%; Pred. No. 0.17;
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 4 ACCGCAAGGTCGCGTATCTTCGCGCAGCGCGGCTTCGAGCATCTGTCGAACC 63
|||
Db 2526 ACCCCGAGGTCCCGGGAACCGCGCGTCCGCGCCCGCCCGCGCTTCGACGAGATC 2585
|||
QY 64 GAACGGGAGTATCGCTCATGGAAGCGCGCTTCGAAAGCGCGTGGACGGTATCGAAGCC 123
|||
Db 2586 TACAAGGCGCGCGCATACAGGGCGGACGTCACCGACGCGGGGCGCCACCAACGCGCGC 2645
|||
QY 124 GTTTCGCGGCGCGCTGCGCTGCGTGCACGTCGACGTTTACGTTGCGCCCGAGTGGCTA 193
|||
Db 2646 GGTACACGGCGGTACGCTGACGGCGCGGCATACCGCGCGCGGTACGCGCGTTCGCTG 2705
|||
QY 184 GATCGCTGAACCGC 199
|||
Db 2706 GACGCGTCGAACGTC 2721
|||

RESULT 14

US-08-152-483B-2
; Sequence 2, Application US/08152483B
; Patent No. 5529909
; GENERAL INFORMATION:
; APPLICANT: Della-Gioppa, Guy
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: TYROSINASE-ACTIVATOR
; TITLE OF INVENTION: PROTEIN FUSION ENZYME
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 2730 Sand Hill Road
; CITY: Menlo Park
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94025

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,483B
; FILING DATE: No. 5529909ember 12, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 7/857,602
; FILING DATE: March 30, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 923,692
; FILING DATE: July 31, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 600,244
; FILING DATE: October 22, 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 641,617
; FILING DATE: January 16, 1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 737,899
; FILING DATE: July 26, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: BIOG-20240/8129-040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4294

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
DESCRIPTION: NO
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM:
IMMEDIATE SOURCE:
CLONE:
FEATURE:
US-08-152-483B-2

Query Match 11.9%; Score 39.2; DB 1; Length 4294;
Best Local Similarity 50.0%; Pred. No. 0.17; 98; Indels 0; Gaps 0;
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 4 ACCGCAAGGTCGCGTATCTTCGCGCAGCGCGGCTTCGAGCATCTGTCGAACC 63
|||
Db 3336 ACCCCGAGGTCCCGGGAACCGCGCGTCCGCGCCCGCCCGCTTCGACGAGATC 3395
|||
QY 64 GAAGCGGGAGTATCGCTCATGGAAGCGCGCGCTTCGAAAGCGCGTGGACGGTATCGAAGCC 123
|||
Db 3396 TACAAGGCGCGCGGATACAGGGCGGCGGTACCGGACGCGGGGCGCCACCAACGCGCGC 3455
|||
QY 124 GTTTCGCGGCGCGCTGCGCTGCGCAGCGTGCACGTTTACGTTGCGCCCGAGTGGCTA 183
|||
Db 3456 GGTACACGGCGGTGACGGTACCGCGCGGCGGCGCATCACGCGCGCGGTACGCGCGTTCGCTG 3515
|||
QY 184 GATCGCTGAACCGC 199
|||
Db 3516 GACGCGTCGAACGTC 3531
|||

RESULT 15

US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: FB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:


```
;
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match          11.6%; Score 38.4; DB 4; Length 1830121;
Best Local Similarity 51.3%; Pred. No. 0.81;
Matches 115; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

QY      45 GAGCATCTCGTCAAAACCGAAGCGGAGTATCGCTCATGGAAGCGGCCCTTCTGAACGGC 105
Db      394522 GAAGGTATCGTCTGATGCTGCAACAGGAGATAACTTATTAGAAGTTGCTCACAATGCG 394463

QY      106 GTGGACGGTATCGAAGCGGTTTGGCGGGCGCTGTGCTGGCCACGTGCCACGTTTAC 165
Db      394462 GGAGTAGAAATTTCATCACGCTTGTATGTTCTTTGTGCTTGTAACCTTGCCACGT---G 394406

QY      166 GTTGCCCCCGAGTGGCTAGATCGCTGAAACCGCGAGTGAGACCGAAGACGAAATGCTC 225
Db      394405 ATTGTTCTGTAGGGGTTTGATTCTTTAAATGAACGAGTGATCAAGAAGAGATATGCTA 394346

QY      226 GATTGGTAGCGGACGTCGCGCGCATTCGCGGCTGTCTGCCA 269
Db      394345 GATAGGCTTGGGATTGAATGATAGCCGTTTATCTTGTC 394302
```

Search completed: June 19, 2004, 07:23:01
Job time : 46.2924 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2004, 06:08:35 ; Search time 156.115 Seconds
(without alignments)
9683.657 Million cell updates/sec

Title: US-10-658-691-2
Perfect score: 330
Sequence: 1 atgacgcgaaggtccggt.....tgccgaagcacagtcagta 330

Scoring table: Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	330	100.0	330	17	US-10-658-691-2
2	79.4	24.1	318	15	US-10-330-662-4
3	79.4	24.1	318	15	US-10-330-661-4
4	65.6	19.9	306	9	US-09-738-626-610
5	65.6	19.9	318	9	US-09-738-626-612
6	65.6	19.9	3309400	9	US-09-738-626-61
7	60.4	18.3	250	10	US-09-535-459-815
8	60.4	18.3	458	10	US-09-918-995-25889
9	60.4	18.3	473	10	US-09-918-995-20620
10	60.4	18.3	554	10	US-09-535-459-431
11	60.4	18.3	958	16	US-10-216-464-12
12	60.4	18.3	958	16	US-10-216-464-19
13	58.8	17.8	294	10	US-09-535-459-430
14	58.2	17.6	413	9	US-09-960-352-7903

RESULT 1
US-10-658-691-2
; Sequence 2, Application US/10658691
; Publication No. US20040110258A1
; GENERAL INFORMATION:
; APPLICANT: Kayser, Kevin J.
; APPLICANT: Kilbane, John J.
; TITLE OF INVENTION: Method for Metabolizing Carbazole in Petroleum
; FILE REFERENCE: GTI-1512
; CURRENT APPLICATION NUMBER: US/10/658,691
; CURRENT FILING DATE: 2003-09-09
; PRIOR APPLICATION NUMBER: US 60/409,562
; PRIOR FILING DATE: 2002-09-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 2
; LENGTH: 330
; TYPE: DNA
; ORGANISM: Sphingomonas sp.
US-10-658-691-2

Query Match 100.0%; Score 330; DB 17; Length 330;
Best Local Similarity 100.0%; Pred. No. 2.2e-91;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGACGCGCAAGGTCCGCGTGTCTTCGCGAGCGCGCGGTTCGAGCATCTGTCGAA	60
Db	1	ATGACGCGCAAGGTCCGCGTGTCTTCGCGAGCGCGCGGTTCGAGCATCTGTCGAA <td>60</td>	60
Qy	61	ACCGAAGCGGAGTATCGCTCATGGAAGCGCGCGGTTCGAA	120
Db	61	ACCGAAGCGGAGTATCGCTCATGGAAGCGCGCGGTTCGAA	120
Qy	121	GCGTTTTCGCGGGCGCGTGTGCTCGCCACGTCACGTTTACGTTGCCCCGAGTGG	180
Db	121	GCGTTTTCGCGGGCGCGTGTGCTCGCCACGTCACGTTTACGTTGCCCCGAGTGG	180

Sequence 14265, A
Sequence 432, App
Sequence 429, App
Sequence 819, App
Sequence 427, App
Sequence 96, Appli
Sequence 6479, App
Sequence 103, App
Sequence 3892, App
Sequence 1, Appli
Sequence 2826, App
Sequence 34363, A
Sequence 1, Appli
Sequence 812, App
Sequence 816, App
Sequence 822, App
Sequence 42250, A
Sequence 1511, App
Sequence 5103, App
Sequence 26053, A
Sequence 804, App
Sequence 24676, A
Sequence 26383, A
Sequence 23574, A
Sequence 33497, A
Sequence 21728, A
Sequence 31012, A
Sequence 1, Appli
Sequence 1, Appli
Sequence 332, App
Sequence 38, Appli

ALIGNMENTS

US-10-658-691-2
; Sequence 2, Application US/10658691
; Publication No. US20040110258A1
; GENERAL INFORMATION:
; APPLICANT: Kayser, Kevin J.
; APPLICANT: Kilbane, John J.
; TITLE OF INVENTION: Method for Metabolizing Carbazole in Petroleum
; FILE REFERENCE: GTI-1512
; CURRENT APPLICATION NUMBER: US/10/658,691
; CURRENT FILING DATE: 2003-09-09
; PRIOR APPLICATION NUMBER: US 60/409,562
; PRIOR FILING DATE: 2002-09-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 2
; LENGTH: 330
; TYPE: DNA
; ORGANISM: Sphingomonas sp.
US-10-658-691-2

Query Match 100.0%; Score 330; DB 17; Length 330;
Best Local Similarity 100.0%; Pred. No. 2.2e-91;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGACGCGCAAGGTCCGCGTGTCTTCGCGAGCGCGCGGTTCGAGCATCTGTCGAA	60
Db	1	ATGACGCGCAAGGTCCGCGTGTCTTCGCGAGCGCGCGGTTCGAGCATCTGTCGAA <td>60</td>	60
Qy	61	ACCGAAGCGGAGTATCGCTCATGGAAGCGCGCGGTTCGAA	120
Db	61	ACCGAAGCGGAGTATCGCTCATGGAAGCGCGCGGTTCGAA	120
Qy	121	GCGTTTTCGCGGGCGCGTGTGCTCGCCACGTCACGTTTACGTTGCCCCGAGTGG	180
Db	121	GCGTTTTCGCGGGCGCGTGTGCTCGCCACGTCACGTTTACGTTGCCCCGAGTGG	180

QY 181 CTAGATGCGGTGAACCGCGAGTGAGACCGAAGACGAATGCTCGATTGCGTAGCGGAA 240
| | | | |
Db 181 CTAGATGCGGTGAACCGCGAGTGAGACCGAAGACGAATGCTCGATTGCGTAGCGGAA 240
| | | | |
QY 241 CGTGCCTGCGATTCGCGGTGCTCTCCAGATCCGCTTACCGACTGCTCGACGGCTG 300
| | | | |
Db 241 CGTGCCTGCGATTCGCGGTGCTCTCCAGATCCGCTTACCGACTGCTCGACGGCTG 300
| | | | |
QY 301 ACCTGGAATCGCGAAGGCACAGTCA 330
| | | | |
Db 301 ACCTGGAATCGCGAAGGCACAGTCA 330
| | | | |

RESULT 2

US-10-330-662-4
; Sequence 4, Application US/10330662
; Publication No. US20030115626A1
; GENERAL INFORMATION:
; APPLICANT: Weeks, Donald
; APPLICANT: Wang, Xiao-Zhuo
; APPLICANT: Herman, Patricia
; TITLE OF INVENTION: Methods and Materials for Making and Using Transgenic Dicamba-Deg
; FILE REFERENCE: 3553-18-1
; CURRENT APPLICATION NUMBER: US/10/330,662
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: US/09/797,238
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/055,145
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/042,666
; PRIOR FILING DATE: 1997-04-04
; PRIOR APPLICATION NUMBER: 60/042,941
; PRIOR FILING DATE: 1997-04-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Pseudomonas maltophilia DI-6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(318)
US-10-330-662-4

Query Match 24.1%; Score 79.4; DB 15; Length 318;

Best Local Similarity 55.7%; Pred. No. 1.4e-14;
Matches 152; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 55 GTCGAAACCGAAGCGGAGTATCGCTCATGGAAGCGCGCTTCTGAACCGCGTGGACGGT 114
| | | | |
Db 46 GTCGAGCGAGTGAAGCGCGCACCCCTGATGGAAGTATCGCGACAGCGGTTTGGACGAA 105
| | | | |
QY 115 ATCGAAGCGTTTGGCGGGCGCTGCTGCTGCGCCACGCTGCCACGTTTACGTTGGCCCC 174
| | | | |
Db 106 CTCCTGGCGTTTGGCGGGCGCTGCTGCTGCGCGACTGCGCACATCGACCCG 165
| | | | |
QY 175 GAGTGGCTAGATGCGCTGAAACCGCGAGTGAGACCGAAGACGAAATGCTCGATTGCGTA 234
| | | | |
Db 166 GCCTTCATGACAGCTGCCGGAGATGAGCGAAGACGAGACGACCTGCTCGACAGCTCG 225
| | | | |
QY 235 GCGGAAGCTGCGCGCATTCGCGGCTGCTGCTGCGCGACTGCGCACATCGACCCG 294
| | | | |
Db 226 GACCACCGCAACGAGTACTCGCTGCTGCGCGACTGCGCACATCGACCCG 165
| | | | |
QY 295 GGCTGACCTTGAACCTGCGAAGGCACAGTCA 327
| | | | |
Db 286 GGCATCAAGTGACGATCGCGCAGGAGACTGA 318
| | | | |

RESULT 3

US-10-330-661-4
; Sequence 4, Application US/10330661

; Publication No. US20030135879A1
; GENERAL INFORMATION:
; APPLICANT: Weeks, Donald
; APPLICANT: Wang, Xiao-Zhuo
; APPLICANT: Herman, Patricia
; TITLE OF INVENTION: Methods and Materials for Making and Using Transgenic Dicamba-Deg
; FILE REFERENCE: 3553-18-1
; CURRENT APPLICATION NUMBER: US/10/330,661
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: US/09/797,238
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/055,145
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/042,666
; PRIOR FILING DATE: 1997-04-04
; PRIOR APPLICATION NUMBER: 60/042,941
; PRIOR FILING DATE: 1997-04-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Pseudomonas maltophilia DI-6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(318)
US-10-330-661-4

Query Match 24.1%; Score 79.4; DB 15; Length 318;

Best Local Similarity 55.7%; Pred. No. 1.4e-14;
Matches 152; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 55 GTCGAAACCGAAGCGGAGTATCGCTCATGGAAGCGCGCTTCTGAACCGCGTGGACGGT 114
| | | | |
Db 46 GTCGAGCGAGTGAAGCGCGCACCCCTGATGGAAGTATCGCGACAGCGGTTTGGACGAA 105
| | | | |
QY 115 ATCGAAGCGTTTGGCGGGCGCTGCTGCTGCGCCACGCTGCCACGTTTACGTTGGCCCC 174
| | | | |
Db 106 CTCCTGGCGTTTGGCGGGCGCTGCTGCTGCGCGACTGCGCACATCGACCCG 165
| | | | |
QY 175 GAGTGGCTAGATGCGCTGAAACCGCGAGTGAGACCGAAGACGAAATGCTCGATTGCGTA 234
| | | | |
Db 166 GCCTTCATGACAGCTGCCGGAGATGAGCGAAGACGAGAACCTGCTCGACAGCTCG 225
| | | | |
QY 235 GCGGAAGCTGCGCGCATTCGCGGCTGCTGCTGCGCGACTGCGCACATCGACCCG 294
| | | | |
Db 226 GACCACCGCAACGAGTACTCGCGTCTCTGTCGAGATTCGGTTCACCGCGGCTCGAA 285
| | | | |
QY 295 GGCTGACCTTGAACCTGCGAAGGCACAGTCA 327
| | | | |
Db 286 GGCATCAAGTGACGATCGCGCAGGAGACTGA 318
| | | | |

RESULT 4

US-09-738-626-610/c
; Sequence 610, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18

```
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 610
; LENGTH: 306
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-610

Query Match      19.9%; Score 65.6; DB 9; Length 306;
Best Local Similarity 58.2%; Pred. No. 2.4e-10;
Matches 135; Conservative 0; Mismatches 94; Indels 3; Gaps 1;

QY 56 TCGAAGCGGAGTATCGCTCATGGAAGCGCGCTTCTGAACGGCGTGGACGGTA 115
Db |||||
286 TCGAGGCGACTGTGGTGATTCAGTAATGGAGACCGCAGTCCGAAACGGAGTGCCTGGAA 227
QY 116 TCGAAGCGGTTTGGCGGGCGCCTGTGCTGCACCGTGCACCGTTTACGTTGGCCCG 175
Db |||||
226 TTGTTGCTGAATGCGGCGGTTCCCTTATCGTGTGCAACTGCCATGTGTTTGAACCTG 167
QY 176 AGTGCTAGATGCGCTGAAACCGCGAGTGAAGCGAAGACGAATGCT---CGATTGG 232
Db |||||
165 CACAGTATGATGCGCTTCCCCCAATGGAGAGATGGAAGATGAAATGCTGTGGGTGCTG 107
QY 233 TAGCGGACGTCGCCCGCATTCGCGGCTGCTCCCGAGATCCGCTTACCGA 284
Db |||||
106 CCGTGGACCGTGAGGATTGCTCCCGTTTGTCTTCCCAATCAAGTCAACCGA 55

RESULT 5
US-09-738-626-612
; Sequence 612, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match      19.9%; Score 65.6; DB 9; Length 3309400;
Best Local Similarity 58.2%; Pred. No. 5.2e-10;
Matches 135; Conservative 0; Mismatches 94; Indels 3; Gaps 1;

QY 56 TCGAAGCGGAGTATCGCTCATGGAAGCGCGCTTCTGAACGGCGTGGACGGTA 115
Db |||||
562917 TCGAGGCGACTTTGTTGATTCAGTAATGGAGACCGCAGTCCGAAACGGAGTGCCTGGAA 562858
QY 116 TCGAAGCGGTTTGGCGGGCGCCTGTGCTGCACCGTGCACCGTTTACGTTGGCCCG 175
Db |||||
562857 TTGTTGCTGAATGCGGCGGTTCCCTTATCGTGTGCAACTGCCATGTGTTTGAACCTG 562798
QY 176 AGTGCTAGATGCGCTGAAACCGCGAGTGAAGCGAAGACGAATGCT---CGATTGG 232
Db |||||
562797 CACAGTATGATGCGCTTCCCCCAATGGAGAGATGGAAGATGAAATGCTGTGGGTGCTG 562738
QY 233 TAGCGGAAACGTCGCCCGCATTCGCGGCTGCTCCCGAGATCCGCTTACCGA 284
Db |||||
562737 CCGTGGACCGTGAGGATTGCTCCCGTTTGTCTTCCCAATCAAGTCAACCGA 562686

RESULT 7
US-09-535-459-815
; Sequence 815, Application US/09535459
```



```

; LENGTH: 958
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (873)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (916)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-216-464-12

Query Match      18.1%; Score 60.4; DB 16; Length 958;
Best Local Similarity 55.1%; Pred. No. 1e-08; 96; Indels 0; Gaps 0
Matches 118; Conservative 0; Mismatches 0

105  CGTGGAGCGTATCGAAGCCGTTTGGCGGGGCGCCTGTGCTCGCGCCACGTGCCACGTTTA 164
291  CGGGGTGGACCTGGAAGGGGSCCTGTGAAGCTCCCTGGCCTGCTCCACCTGCATGTGTA 350

165  CGTTGGCCCCAGTGGCTAGATCGCTGAAACGGCCGAGTGAACGGAGACGAATGCT 224
351  TGTGAGTGAAGACACACCTGGAATCTCTGCTCTCTCCGAGGAGGGAGACGACATGCT 410

225  CGATTGGGTAGCGGAACGTCGCGCGCGCATTCGGGGCTGTCTCCAGAGATCCGCTTTACCGA 284
411  AGACATGCCCCCCCTCCTCCAGAGAACTCGCGGCTGGGCTGCCAGATTGTGCTGACACC 470

285  CCTGCTCGAGCGGCTGACCTGGAACTGCGGAAG 318
471  GGAGCTGGAAGGCGGAATTACCTGCCCAAG 504

Db

```

RESULT 12
US-10-216-464-19
; Sequence 19, Application US/10216464
; Publication No. US20030207285A1

```

1  / GENERAL INFORMATION:
2  / APPLICANT: Rosen et al.
3  / TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
4  / FILE REFERENCE: PTZ19C1
5  / CURRENT APPLICATION NUMBER: US/10/216,464
6  / CURRENT FILING DATE: 2002-08-12
7  / PRIOR APPLICATION NUMBER: 09/764,883
8  / PRIOR FILING DATE: 2001-01-17
9  / PRIOR APPLICATION NUMBER: 60/179,065
10 / PRIOR FILING DATE: 2000-01-31
11 / PRIOR APPLICATION NUMBER: 60/180,628
12 / PRIOR FILING DATE: 2000-02-04
13 / PRIOR APPLICATION NUMBER: 60/214,886
14 / PRIOR FILING DATE: 2000-06-28
15 / PRIOR APPLICATION NUMBER: 60/217,487
16 / PRIOR FILING DATE: 2000-07-11
17 / PRIOR APPLICATION NUMBER: 60/225,758
18 / PRIOR FILING DATE: 2000-08-14
19 / PRIOR APPLICATION NUMBER: 60/220,963
20 / PRIOR FILING DATE: 2000-07-26
21 / PRIOR APPLICATION NUMBER: 60/217,496
22 / PRIOR FILING DATE: 2000-07-11
23 / PRIOR APPLICATION NUMBER: 60/225,447
24 / PRIOR FILING DATE: 2000-08-14
25 / PRIOR APPLICATION NUMBER: 60/218,290
26 / PRIOR FILING DATE: 2000-07-14
27 / Remaining Prior Application data removed - See File Wrapper or PALM.
28 / NUMBER OF SEQ ID NOS: 38
29 / SOFTWARE: PatentIn Ver. 2.0
30 / SEQ ID NO 19
31 / LENGTH: 958
32 / TYPE: DNA
33 / ORGANISM: Homo sapiens
34 / FEATURE:
35 / NAME/KEY: misc feature

```

; LOCATION: (873)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (916)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-216-464-19

Query Match 18.3%; Score 60.4; DB 16; Length 958;
Best Local Similarity 55.1%; Pred. No. 1e-08;
Matches 118; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 105 CGTGGAGGATCGAAGCCGTTTGGGGGGCGCTGTGCTGCGCCACGTCGACGTTTA 164
DB 291 CGGGTGGACCTGGAAGGGCCCTGTGAAGCTCCCTGGCTGTCCACCTGCATGTGA 350
QY 165 CGTTGGCCCCGAGTGTAGATCGCTGAAACCGCGAGTGAGACGAGAAATGCT 224
DB 351 TGTGAGTGAAGACCACTGGATCTCTGCTCTCCGAGGAGAGGGAAGACATGCT 410
QY 225 CGATTGGTAGCGAAGCTGCGCGGCAATTCGGCGCTGTCTGCCAGATCGGCTTACCGA 284
DB 411 AGACATGCCCCCTCTCCAGAGAACTCGCGCTGGGCTGCCAGATTGTGCTGACACC 470
QY 285 CTTGCTCGAGCGGCTGACCCCTGGAATTCGCCGAAG 318
DB 471 GGAGCTGGAAGGAGCGGAATTCACCTTGCCCAAG 504

RESULT 13
US-09-535-459-430
; Sequence 430, Application US/09535459
; Publication No. US20030040615A1
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Deleseane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING ELECTRON TRANSFER MOLECULE
; FILE REFERENCE: PD-1014 CIP
; CURRENT APPLICATION NUMBER: US/09/535,459
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2170
; SOFTWARE: PERL Program
; SEQ ID NO 430
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030040615A1 hu01270499
US-09-535-459-430

Query Match 17.8%; Score 58.8; DB 10; Length 294;
Best Local Similarity 54.7%; Pred. No. 2.9e-08;
Matches 117; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
QY 105 CGTGGAGGATCGAAGCCGTTTGGGGGGCGCTGTGCTGCGCCACGTCGACGTTTA 164
DB 37 CGGGTGGACCTGGAAGGGCCCTGTGAAGCTCCCTGGCTGTCCACCTGCATGTGA 96
QY 165 CGTTGGCCCCGAGTGTAGATCGCTGAAACCGCGAGTGAGACGAGAAATGCT 224
DB 97 TGTGAGTGAAGACCACTGGATCTCTTACCTCTCCGAGGAGAGGGAAGACATGCT 156
QY 225 CGATTGGTAGCGGAAGCTGCGCGCAATTCGGCGCTGTCTGCCAGATCGGCTTACCGA 284
DB 157 AGACATGCCCCCTCTCCAGAGAACTCGCGCTGGGCTGCCAGATTGTGCTGACACC 216
QY 285 CTTGCTCGAGCGGCTGACCCCTGGAATTCGCCGAAG 318

DB 217 GGAGCTGGAAGGAGCGGAATTCACCTGCCCAAG 250
RESULT 14
US-09-960-352-7903
; Sequence 7903, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7903
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 34-LIB3057-016-Q1-K1-A6
US-09-960-352-7903

Query Match 17.6%; Score 58.2; DB 9; Length 413;
Best Local Similarity 55.1%; Pred. No. 4.6e-08;
Matches 114; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 112 GGTATCGAAGCCGTTTGGGGGGCGCTGTGCTGCGCCACGTCGACGTTTACGTTGGC 171
DB 131 GATCTGGAAGGGGCTCGAAGCGTCTTGGCGTGTCCACCTGCCACGTTATGTGAGT 190
QY 172 CCCGAGTGGCTAGATGGCTGAACCGCGAGTGAGACCGAAGAAATGCTCGATTGC 231
DB 191 GAGGACCACTGGACCTTCTGCCGCTCTCTGATGAGAGGGAGGACGACATGCTGGATG 250
QY 232 GTAGCGGAACGTGCGCGCATTCGCGGCTGTCTTCGCGAGATCCGCTTACCGACCTGCTC 291
DB 251 GCCCTCTCTCCAAAGAACTCCCGCTGGCTGCCAGATCGTGTGCTGACGCTGAGCTG 310
QY 292 GACGCGCTGACCTGGAACTGCCGAAG 318
DB 311 GAAGGGCGCGAATTCACCTGCCCAAG 337

RESULT 15
US-09-960-352-14265
; Sequence 14265, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14265
; LENGTH: 443
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 61-LIB3057-009-Q1-K1-H2
US-09-960-352-14265

Query Match 17.6%; Score 58.2; DB 9; Length 443;
Best Local Similarity 55.1%; Pred. No. 4.6e-08;
Matches 114; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 112 GGTATCGAAGCCGTTTGGCGGGCGCCCTGTGCTCGCCACGTCGCCACGTTTACGTTGGC 171
 Db |||||
 QY 130 GATCTGGAAGGGCCCTGCGAAGCGTCCTTGGCGTCTCCACCTGCCACGTTGATGTGAGT 189
 Db |||||
 QY 172 CCCGAGTGGCTAGATCGCTGAAACCGCCGAGTGAGACCGAAGCGAATGCTCGATTGC 231
 Db |||||
 QY 190 GAGGACCACCTGGACCTTCTGCCGCTCTGTGATGAGGGAGGACGACATGCTGGATATG 249
 Db |||||
 QY 232 GTAGCGGAACGTGCGCGCCATTGCGGGCTGTCTCTCCAGATCCGCTTACCGACCTGCTC 291
 Db |||||
 QY 250 GCCCTCTCTCCAAAGAGAACTCCCGGTGGGCTGCCAGATCGTCTGACGCTGAGCTG 309
 Db |||||
 QY 292 GACGCGCTGACCCCTGGAACCTGCGAAG 318
 Db |||||
 QY 310 GAAGGGCGGAATCACCTGCCCAAG 336
 Db |||||

Search completed: June 19, 2004, 09:16:55
 Job time : 171.115 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 19, 2004, 04:04:20 ; Search time 4587.53 Seconds
(without alignments)
10742.394 Million cell updates/sec

Title: US-10-658-691-1
Perfect score: 1137
Sequence: 1 gtggtaaccaacatcaat.....gcattcaggagcagttga 1137

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rod.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sy.*
39: em_hgt_hum.*
40: em_hgt_mus.*
41: em_hgt_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1137	100.0	4154	1	AF442494	AF442494 Spingomo
2	1137	100.0	13651	1	AB095953	AB095953 Spingomo
3	859.2	75.6	1110	1	AB072827	AB072827 Spingomo
4	422.6	37.2	20017	1	AB095952	AB095952 Janthinob
5	419.4	36.9	1161	6	E17292	E17292 gDNA encodi
6	419.4	36.9	1200	6	E17293	E17293 gDNA encodi
7	419.4	36.9	6880	1	AB001723	AB001723 Pseudomon
8	419.4	36.9	6881	1	D89064	D89064 Pseudomonas
9	419.4	36.9	44266	1	AB047548	AB047548 Pseudomon
10	419.4	36.9	199035	1	AB088420	AB088420 Pseudomon
11	281.8	24.8	8096	1	AB094786	AB094786 Pseudomon
12	153.6	13.5	3203	1	PPY12655	Y12655 P.putida ox
13	49	4.3	2000	6	AX655393	AX655393 Sequence
14	48.8	4.3	298450	1	SC039107	AL939107 Streptomy
15	47.4	4.2	125020	9	AF429315	AF429315 Homo sapi
16	47	4.1	125020	9	AF429315	AF429315 Homo sapi
17	46.2	4.1	6564	6	AX069073	AX069073 Sequence
18	46.2	4.1	32870	1	AE007122	AE007122 Mycobacte
19	46.2	4.1	318050	1	EX248344	EX248344 Mycobacte
20	46.2	4.1	348676	15	BX842581	BX842581 Mycobacte
21	44	3.9	5727	1	REU24277	U24277 Rhodococcus
22	44	3.9	210285	1	AY223810	AY223810 Rhodococc
23	43	3.8	344321	1	EX640429	EX640429 Bordetell
24	43	3.8	348624	1	EX640441	EX640441 Bordetell
25	42.8	3.8	2000	6	AX655393	AX655393 Sequence
26	41.6	3.7	5957	1	RERBP41	D32142 Rhodococcus
27	41	3.6	15982	1	AE004779	AE004779 Pseudomon
28	41	3.6	89153	1	NSP561198	AJ561198 Actinomad
29	40.6	3.6	15663	1	AE007046	AE007046 Mycobacte
30	40.6	3.6	291050	1	EX248340	EX248340 Mycobacte
31	40.6	3.6	301617	1	AE016911	AE016911 Chromobac
32	40.6	3.6	347496	1	EX842577	EX842577 Mycobacte
33	40.6	3.6	348171	1	EX640412	EX640412 Bordetell
34	40.2	3.5	14886	1	TAU52917	U52917 Thermus aqu
35	40.2	3.5	349640	1	EX572600	EX572600 Rhodopsu
36	40	3.5	3151	1	AB055506	AB055506 Streptomy
37	40	3.5	4483	1	PPPOB4B	X78823 P.pseudoalci
38	40	3.5	6835	1	AF452376	AF452376 Rhodococc
39	40	3.5	20021	1	AE004730	AE004730 Pseudomon
40	40	3.5	299350	1	AP006572	AP006572 Gloeobact
41	39.8	3.5	606	6	AX196058	AX196058 Sequence
42	39.8	3.5	109519	6	AX195929	AX195929 Sequence
43	39.4	3.5	10177	1	AE004977	AE004977 Halobacte
44	39.4	3.5	304500	1	AP005953	AP005953 Bradyrhiz
45	39.2	3.4	15873	1	AB007115	AB007115 Mycobacte

ALIGNMENTS

RESULT 1
AF442494
LOCUS
DEFINITION Spingomonas sp. GTIN11 carbazole operon, partial sequence.
ACCESSION AF442494
VERSION AF442494.1 GI:17227030
KEYWORDS Spingomonas sp., GTIN11
SOURCE Spingomonas sp. GTIN11
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Spingomonadales;
Spingomonadaceae; Spingomonas.
REFERENCE 1 (bases 1 to 4154)
AUTHORS Kilbane, J. J., Daram, A., Abbasian, J. and Kayser, K. J.
TITLE Isolation and characterization of Spingomonas sp. GTIN11 capable
of carbazole metabolism in petroleum

JOURNAL Biochem. Biophys. Res. Commun. 297 (2), 242-248 (2002)
MEDLINE 22222130
PUBMED 12237109
REFERENCE 2 (bases 1 to 4154)
AUTHORS Kayser, K.J., Daram, A., Abassian, J. and Kilbane, J.J. III.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2001) Biotechnology, Gas Technology Institute,
1700 South Mount Prospect Road, Des Plaines, IL 60018, USA

FEATURES
source
1..4154
/organism="Sphingomonas sp. GTIN11"
/mol_type="genomic DNA"
/db_xref="taxon:174635"
<1..>4154
/gene="carbazole operon"
650..1786
/gene="carAa"
650..1786
/gene="carAa"
/codon_start=1
/transl_table=11
/product="carbazole dioxygenase"
/protein_id="AAL37976.1"
/db_xref="GI:17227031"
/translations="MANQSIARRTKVEPIRAKLGFRNHWYVRLASEGTPV
PVKLGEKILNVRGKVAIQDCLRHGVTLSDRVECYSKNTISCWYHGWTYRLWDDG
RLVDILTNPQSVQIGRRALKTPVEEAKGLIFVYDVGSEPTLIEDVPPGLDENRAI
HGQRLVAVNWRGLAGNDFGADHFIHKNSILVKGNDIILPLGFAPGPDQDLTRSEVA
AGKPGVYDLGSHSVFVPEMIEGKPAIHGNIKSVAISISIMLPGLVKVPEPDP
ELTQFEWYVVDTSLSHYFTLGLKVVTSEKAADSPEPEHKEWVGLALNGFNDDDDIMA
RESMEPPYADDRGWSBEILFEPPDRAILIEWRGLASQHNRCIQEAR"
1735..2067
/gene="carBa"
1735..2067
/gene="carBa"
/notes="meta cleavage enzyme subunit"
/codon_start=1
/transl_table=11
/product="2-aminobiphenyl-2,3-diol 1,2-dioxygenase"
/protein_id="AAL37977.1"
/db_xref="GI:17227032"
/translations="MAGACQSQSHSGSLTMGTASYPVNRMLQELFTNPGNVELFR
ADREALYERGLSSAQRAALDEGGFALTAUGLHPVLQMHFMLTNPMAPDFVSKAY
RKMVDRNG"
2060..2863
/gene="carBb"
2060..2863
/gene="carBb"
/notes="meta cleavage enzyme catalytic subunit"
/codon_start=1
/transl_table=11
/product="2-aminobiphenyl-2,3-diol 1,2-dioxygenase"
/protein_id="AAL37978.1"
/db_xref="GI:17227033"
/translations="MAEIVGFATSHVLFSPDGAQALRVVEGNEIGRRVRLLRPD
LLVIGSDHFNITRLQPPFTVGSVDFTPLGMDIEORFPAGNRAFAESICARAAD
RFDLAQEERLPHGVNVPVPMFTDPDRIPVPIVYVNNPTPPYTAAGALQGLDIAVE
AVGLDLPSHLRVVVAATGSLSHWINIPGHGVNAEFDRRVIAAFQSGDMRWLRADITE
TLKKNAGSGLEIVNWMAAATLPGRAEKIYVEPQWNTGNGGAIIV"
2906..3730
/gene="carC"
2906..3730
/gene="carC"
/codon_start=1
/transl_table=11
/product="meta cleavage compound hydrolase"
/protein_id="AAL37979.1"
/db_xref="GI:17227034"
/translations="MVLAAVAGRSVTVRGNETYHHEQSGDVVVLVHGGAGADSMGN
WRGMPVLADRYRIVADMLGFGRTAKPADPPVFSQAARTDHLAGLGLSNVALV
GNSNGSASALGAVRPLVRLKMSAGLSKIDPALEPLVGYDFTREGMIRLVRA
LITDNDFOIDAMIDYRALGSDVPETRAYSATMOWIRDQGLVYEDDYIRRTAPTLL
VNGKLDKRVPLANAKFELIIGPSWGVIMPDCGHWAMIEHPVDFAITTAFFIEAAQ"

gene
3771..4100
/gene="carAc"
3771..4100
/gene="carAc"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="AAL37980.1"
/db_xref="GI:17227035"
/translations="MTAKVPIFRAAGGPEHLVETAGVSLMEAVLNGVDGTEAVCG
GACATCHVYVGPPEHLDAKKPSETEDMLDCVAERAPHSRLSCQIRLTDLDDGLTL
ELPKAQS"

ORIGIN
Query Match 100.0%; Score 1137; DB 1; Length 4154;
Best Local Similarity 100.0%; Pred. No. 7.7e-250;
Matches 1137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGCTAACCAACCATCAATCGCGGAGCGGAGAACCAAGGTTTGGGAGCCTTATATCCGT 60
DB 650 GTGGCTAACCAACCATCAATCGCGGAGCGGAGAACCAAGGTTTGGGAGCCTTATATCCGT 709
QY 61 GCGAAACTCGGGTTCGGAACCATTTGGTATCCGCTCGCGAGCGAAATCGCCGAA 120
DB 710 GCGAAACTCGGGTTCGGAACCATTTGGTATCCGCTCGCGAGCGAAATCGCCGAA 769
QY 121 GGTACTCCGTTCCGTCACAGCTCCTGGAGAGAGAAATCTGTCAATCGCGTGGCGGC 180
DB 770 GGTACTCCGTTCCGTCACAGCTCCTGGAGAGAGAAATCTGTCAATCGCGTGGCGGC 829
QY 181 AAGTCTATGCGATCCAGGACAGGTGCTGCATCGCGGTGTAAAGCTTTCGACCGGGTC 240
DB 830 AAGTCTATGCGATCCAGGACAGGTGCTGCATCGCGGTGTAAAGCTTTCGACCGGGTC 889
QY 241 GAGTCTATTCAGAAACACCATATCTGTGTGTATCACGGCTGGAACATATCGTGGGAC 300
DB 890 GAGTCTATTCAGAAACACCATATCTGTGTGTATCACGGCTGGAACATATCGTGGGAC 949
QY 301 GATGGCCCTCGTGCATATCTCACAACCCCGGAGTGTGCAGATCGCGCGCGCGCT 360
DB 950 GATGGCCCTCGTGCATATCTCACAACCCCGGAGTGTGCAGATCGCGCGCGCGCT 1009
QY 361 TTGAAGAGCTTCCCGTTGAAGAGGCCAAAGTCTTATCTTCTGTTACGTAGGCGACGGC 420
DB 1010 TTGAAGAGCTTCCCGTTGAAGAGGCCAAAGTCTTATCTTCTGTTACGTAGGCGACGGC 1069
QY 421 GAACCAAGCCGCTTATCGAAGATGTGCGCCCGCGCTTCTTGTATGAACCGCGCAT 480
DB 1070 GAACCAAGCCGCTTATCGAAGATGTGCGCCCGCGCTTCTTGTATGAACCGCGCAT 1129
QY 481 CACGCCCAACATCGCTCGTGGCTCGAACTGGCGCTTGGTGCAGAAACGGCTTTGAT 540
DB 1130 CACGCCCAACATCGCTCGTGGCTCGAACTGGCGCTTGGTGCAGAAACGGCTTTGAT 1189
QY 541 GCGGGGACGCTTCAATTCACAGAAATCGATTCGTGAAGGCGACGATATCATCTG 600
DB 1190 GCGGGGACGCTTCAATTCACAGAAATCGATTCGTGAAGGCGACGATATCATCTG 1249
QY 601 CGCTTGGCTTTCGCGCTCGCGATCCGACAGCTTACCGCTTCGAGGTTGCTGCGGGC 660
DB 1250 CGCTTGGCTTTCGCGCTCGCGATCCGACAGCTTACCGCTTCGAGGTTGCTGCGGGC 1309
QY 661 AAGCCCAAGGTTTACGATCTGTTGGCGAGCATTCGCTGCGCGTTTCGAAAGGATG 720
DB 1310 AAGCCCAAGGTTTACGATCTGTTGGCGAGCATTCGCTGCGCGTTTCGAAAGGATG 1369
QY 721 ATCGAAGCAACCTGCAATCCATGGAACATTCGAGAGCGCGCTCGCCATCAGGATA 780
DB 1370 ATCGAAGCAACCTGCAATCCATGGAACATTCGAGAGCGCGCTCGCCATCAGGATA 1429
QY 781 TCGATCTGGCTGCCGGGCGTACTCAAGTGCAGACCGTGGCGGATCCCGAGCTCACGCG 840
DB 1430 TCGATCTGGCTGCCGGGCGTACTCAAGTGCAGACCGTGGCGGATCCCGAGCTCACGCG 1489

```
QY      841  TTCGAATGCTAGTCCGGTTCAGTCCAGACCGACCTCTACTTCCAGACGCTGGGAAA  900
Db      1490  TTCGAATGCTAGTCCGGTTCAGTCCAGACCGACCTCTACTTCCAGACGCTGGGAAA  1549
QY      901  GTCGTGACGTCAAGAGAGCGGACGACCTCTTCGAGCGAGAAATCCAGAAAATGGGTA  960
Db      1550  GTCGTGACGTCAAGAGAGCGGACGACCTCTTCGAGCGAGAAATCCAGAAAATGGGTA  1609
QY      961  GSCCTCGCGCTTAAACGGCTTCAATCATGACGACATCATGGCAGCGATCATGAGCGCG  1020
Db      1610  GSCCTCGCGCTTAAACGGCTTCAATCATGACGACATCATGGCAGCGATCATGAGCGCG  1669
QY      1021  TTCTACGCTGATGATCGCGGTTGGTCCGAGAAATCTCTTTCGAGCCGAGCGCGCAATC  1080
Db      1670  TTCTACGCTGATGATCGCGGTTGGTCCGAGAAATCTCTTTCGAGCCGAGCGCGCAATC  1729
QY      1081  ATCGAGTGGCGGGGCTTCCAGTCAGCACATCTCGGCATTCAGAGACGAGCTTGA  1137
Db      1730  ATCGAGTGGCGGGGCTTCCAGTCAGCACATTCGCGGATTCAGAGACGAGCTTGA  1786

RESULT 2
LOCUS   AB095953                      13651 bp      DNA      linear      BCT 04-FEB-2003
DEFINITION   Spingomonas sp. KAI car operons, complete cds.
ACCESSION   AB095953
VERSION     AB095953.1  GI:28201219
KEYWORDS
SOURCE      Spingomonas sp. KAI
            Spingomonas sp. KAI
            Bacteria; Proteobacteria; Alphaproteobacteria; Spingomonadales;
            Spingomonadaceae; Spingomonas.
REFERENCE   1
AUTHORS    Inoue, K., Nojiri, H., Nakai, S., Endoh, T., Urata, M., Ashikawa, Y.,
            Saiki, Y., Yoshida, T., Habe, H., and Omori, T.
TITLE      Divergent structures of carbazole-degrading car operons isolated
            from several gram-negative bacteria
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 13651)
AUTHORS    Nojiri, H. and Inoue, K.
TITLE      Direct Submission
JOURNAL    Submitted (11-NOV-2002) Hideaki Nojiri, Biotechnology Research
            Center, The University of Tokyo; 1-1-1, Yayoi, Bunkyo-ku, Tokyo
            113-8657, Japan (E-mail:anojiri@mail.ecc.u-tokyo.ac.jp,
            Tel.81-3-5841-3067, Fax:81-3-5841-8030)
FEATURES   Location/Qualifiers
            1..13651
                /organism="Spingomonas sp. KAI"
                /mol_type="genomic DNA"
                /strain="KAI"
                /db_xref="taxon:164608"
                complement(127..2199)
                /gene="ORF10"
                complement(127..2199)
                /gene="ORF10"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAC56754.1"
                /db_xref="GI:28201220"
                /translation="MISEKIGPOHLERKALYVROSSVHQVHNRESSTLOYAMRGL
                TANGWSHIETDIDLGSAAGVTRAGDFRMVAEYCLGKGVAAAREVSRFARNRDW
                QOLIEKRVVDVTLIDQETVYAPQONDRLILGLKSLNEVELDLRLQRLSARYEKA
                RQELVVAAPGVFVAGDRIEKDPDRRQVEALVDFKVAELGSAQALLWFIEHGLD
                LPAKRNNGDVWRPSYATIRMIENPIYGGAAYGKSRATGYDGTAMRSRERKAR
                DWLTYLPGSHGYSVWSESDIRKXNSDNPVTSOHGAPKHGDALLAGLIRCRGCR
                KURTAGCHDIPRYSCWRGLDNGEPCFATGRLRVDDALERALLQVLEFGTAAS
                VAEQAQRDQDVRLNRDLAEARYADAFRDYDAADPQNRVLVAELETRMNRAL
                TEAGEVEARIVAHDASTAPALPSLKDIDGLASLEAVNAPQSDARLKCIIVRTLQ
                EVVADIDHDASIVILLIHWVGGVHTDLRLPRRKGQNSTGADIIAAREVLVLIANDI
                LIAGILNRNLVTHGNRNRTRERVTLARSHHRIPVFRTVADGPAPWLNLQAARHIGV
```

```
gene
CDS
ASTKRIAAEAGEIKGIHPLPEGPWIFCRTELDGSAAHHLAKRARQNPKYPTGSHPDQ
QTLFSSTT"
2502..2852
/gene="ORF9"
2502..2852
/gene="ORF9"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56755.1"
/db_xref="GI:28201221"
/translation="MGRTHPCQAGARSVEHPRNLEPPVRSRSRLATACINRTGSFD
LMDVMPASPGMAIPENFSNFGPVGASPCCTTSARTAVLAIGHRQKRHRHIRP
PVPLRSTSVRIWR"
3162..3785
/gene="ORF8"
3162..3785
/gene="ORF8"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56756.1"
/db_xref="GI:28201222"
/translation="MKYAKASERFHEYGITKSIKTPYNTVHTLOEVNDGLAGHY
FMVDKDGIALYQSDSDVLHTPKPKTPDALKMAREYEDENFPSSMKYNTLAKFVG
QGLYRDAADFHQATRELHYHCVLLVCTFYTPVHNHGLFRLSQANLIDRLRLMYWPEDN
RKQVMFEKLQAYVKARYSKHYRISEBELTWLGEQVBEIGRVVHAV"
3862..4188
/gene="ORF7"
3862..4188
/gene="ORF7"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56757.1"
/db_xref="GI:28201223"
/translation="MQRDSDSPRONGRSATKNRDLIAAFVSCANAGFRKRREAGSGS
NFGFFQQRRIIDIYAEVSDCLDVGMAEQDLYGAEIACGFVNRRLRSHRVRAITFP
TVETNR"
complement(4409..5089)
/gene="car"
complement(4409..5089)
/gene="car"
/codon_start=1
/transl_table=11
/product="transcriptional regulator of car operon"
/protein_id="BAC56758.1"
/db_xref="GI:28201224"
/translation="MISTMTSLDKESPAHRAAATIRGAILEARFAPGQRLVSELS
SLTGLGPGVREALRLAAGLVAIRNGAVARAKAMVVEVFELEGLAARR
AALRAATSPHLESIERALDEESLRNVDGLILMEANEHLHLQVVDAAESPVIERLV
QMQLPELSSIFFRLMSFTIWRHSREDHIGLINTILKDGGEAASQRAHVRYTAALFA
DLPDELIR"
5192..6328
/gene="carAa"
5192..6328
/gene="carAa"
/codon_start=1
/transl_table=11
/product="terminal oxygenase component of car
1, 9a-dioxygenase"
/protein_id="BAC56759.1"
/db_xref="GI:28201225"
/translation="MANQPSIAERRTKWPEYTRAKLGFNRHWYPVRLASEIAEGTPV
PVLLGKILINRVGKVVAIDRCLHRGVTLSDRVEYKNTKISWYHGTVYRWDDG
RLVDLTNPQSVQIGRALKATPVEEAKGLIYVVGDEGPTPLIEDVPGLDENRAI
AGKRLVYNWRNLGAENGDFGAVFIRKNSILYKGNLILPLGFAQDPPQLTRSEVA
HQPRLGYVDLGEHSPVFEPMIEGPAHNGTSKRVAISISIMLPGVZKVEFPDP
ELTQFEYVYVDETSILYFQTLGKVTSEADSFEREFHEKRVGLANGFNDDIMA
RESMEPYADDRGWSBELPEPDRAIENRGLASQHRNGIQEAR"
```

```
6328..6609
/gene="carBa"
CDS
/transl_table=11
/product="subunit of meta cleavage enzyme"
/protein_id="BAC56760.1"
/db_xref="GI:28201226"
/translation="MGTTASYPVNRLMQELFTNPGNVELFRADREALYRYGLSSAQR
AALDEGGGALTAVGLHPVLQWHFHLTNWPDPFVSVAVRKWDNRG"
6602..7405
/gene="carBb"
CDS
/transl_table=11
/product="subunit of meta cleavage enzyme"
/protein_id="BAC56761.1"
/db_xref="GI:28201227"
/translation="MAEIVGFVATSHVLFGPSDDQAALRVVVEGMEIGRVRLLRPD
LLIVGSHLENIITRIQPPFTVGVSDTFTPLGMDIEQRFAGNAPAFSLCARAAD
RFDLAQGELEPDRHGVWVPLMFIIDPDGRIPVPIYVYNNMTFPTAARAAQLGDIIVAE
AVGLDPLSHLRVAVATGGLSHWINIPGHGEVNAEFORRVIAAFQSGDMRWLRADTE
TLKNAGNGGLEIIVNWMMAATLPGRAEKIYVBPMPQWMTGMGIAIV"
7448..8272
/gene="carC"
CDS
/transl_table=11
/product="meta cleavage compound hydrolase"
/protein_id="BAC56762.1"
/db_xref="GI:28201228"
/translation="MWLAATAGRSVTVRGMEITYHEQSGDVVVLVHGGAGADSMGN
WGVNPLVADRIRVAVEMLGFGRKAPDFVFSQARTDHLAGFLDALGSLNVALV
GNSMGAGALVAVERPGLVRLKVLMSGAGLVSKIDPALEPLVGYDFTREGMIELVRA
LTNDFQDDAMIDYRVALSDVPETTRAYSATMQWIDQGLGYEDDYIRITAPTII
VNGKLDKVPVLANAYKFLLELIGPSMGVIMPCDGHAMIEHPVDVPTAARFIEAAQ"
8313..8642
/gene="carAc"
CDS
/transl_table=11
/product="ferredoxin component of CAR 1,9a-dioxygenase"
/protein_id="BAC56763.1"
/db_xref="GI:28201229"
/translation="WNAKRVIFRAAGGFHELVETEAGVSLMEAAVLNGVDGIEAVCG
GACACATCHVYVGPBWLDAKFPFSETEDEMLDCVAERAPHSRLSCQINLTDLLGLTL
ELPKAQS"
8684..10999
/gene="ORF11"
CDS
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56764.1"
/db_xref="GI:28201230"
/translation="WCKLIMSGALAVLASAPAAQOTDGVEDVASELEIITVTAER
VSENLQDSPAIVASARGENAIRAINVSDIGKTPSLSPDASAAAGSSSVTVFLR
GTGQDFNMVDPGVGILYDGYISRSVGALLETRDVESIQVLRGPQOTLFGKNTIG
AVITSRPDEQNGNIEATGDSYINRDIYDINRINRVSEDFRFVYSGALTRDYVER
LVDGKQNGNSYGRFAEMDADPFDNLLISADYTRREAVATLLDNNENAIFPL
FHNVFLNAPGLPPTSPSCYNQMLTGDKNTESGNRSVSDVWVGSLNSWDA
GVTIKTSITAYRELSAFAVIDQSPLAIQRTENDYQNSQSFQLOGDGFGEKLYN
TLGLVYHGGGTDVNSLFTAPTDFSSGSKVNDVYAAFAQFIHATDNLNLTCGRYT
SEDKFSPDOEVLVHBAQAPLPSPFTBPRLTNVPTVMAAGSPQLRVGDIIVPAV
QRTTSANFTPEVTLDYKFTDMLGYFTFSGFKSGGRTQVFPPEPAPEFEFAR
VLEIGVVELFNRLNSLFDFTIKYISDLQIIVNDGPAPKRNAGKIRIFESEAQL
VVTDAFRITGAVSYLDABYREVDAARAPVTVDLSALANTPKWTANAGITVEPDEAGR
```

```
VVLQDWSYRSATYKDAINTPSLKQDGYSVFNMSATFDAAQGSVGTGITNLTDKRYI
ISGYDLNFIPIGSTYGVYRPRWYKLGYSF"
complement(11435..12808)
/gene="ORF12"
complement(11435..12808)

Query Match 100.0%; Score 1137; DB 1; Length 13651;
Best Local Similarity 100.0%; Pred. No. 7.6e-250;
Matches 1137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGCTAAACCAACCATCAATCGCGAGCGCAGAAACCAAGGTTTGGAGCCTTATATCCGT 60
DB 5192 GTGGCTAACCAACCATCAATCGCGAGCGCAGAAACCAAGGTTTGGAGCCTTATATCCGT 5251
QY 61 GCGAAACTCGGGTTCGAAACCAATTTGGTATCCCGTTCGCTCGCGAGCGAAATCGCGAA 120
DB 5252 GCGAAACTCGGGTTCGAAACCAATTTGGTATCCCGTTCGCTCGCGAGCGAAATCGCGAA 5311
QY 121 GGTACTCCCGTTCGCGTCAAGCTCCTGGGAGAGAGATTCTCTCAATCGCGTGGCGGC 180
DB 5312 GGTACTCCCGTTCGCGTCAAGCTCCTGGGAGAGAGATTCTCTCAATCGCGTGGCGGC 5371
QY 181 AAGGTCATCGCATCCAGGACAGGTGCTGCATCGCGGTGTAAACGCTTTCGACCGGGTC 240
DB 5372 AAGGTCATCGCATCCAGGACAGGTGCTGCATCGCGGTGTAAACGCTTTCGACCGGGTC 5431
QY 241 GAGTGTATTCNAAGAACCATATCTCTGCTGATCACGGCTGGACATATCGCTGGAC 300
DB 5432 GAGTGTATTCNAAGAACCATATCTCTGCTGATCACGGCTGGACATATCGCTGGAC 5491
QY 301 GATGGCGGCTCGTGATATCTCTCAAAACCCCGGAGTGTGCAGATCGCGCGCGCGT 360
DB 5492 GATGGCGGCTCGTGATATCTCTCAAAACCCCGGAGTGTGCAGATCGCGCGCGCGT 5551
QY 361 TTGAAGACGTTCCCGGTTTGAAGAGCGCAAGGCTCTTATCTTGTTCAGTACGAGCGCGC 420
DB 5552 TTGAAGACGTTCCCGGTTTGAAGAGCGCAAGGCTCTTATCTTGTTCAGTACGAGCGCGC 5611
QY 421 GAACCAACGCGCTTATCGAAGATGTGCGCGCGCGCTTCTCTTGATGAAACCGCGCAIT 480
DB 5612 GAACCAACGCGCTTATCGAAGATGTGCGCGCGCGCTTCTCTTGATGAAACCGCGCAIT 5671
QY 481 CAGCGCAACATCGGCTCGTGCGCTGAACTGCGGCTTGGTTCGCGAAGCGCTTGTAT 540
DB 5672 CAGCGCAACATCGGCTCGTGCGCTGAACTGCGGCTTGGTTCGCGAAGCGCTTGTAT 5731
QY 541 GCGGGGCACTCTTCAATTCAGAAATTCGATCCTCGTGAAGGCAACGATATCATCTCG 600
DB 5732 GCGGGGCACTCTTCAATTCAGAAATTCGATCCTCGTGAAGGCAACGATATCATCTCG 5791
QY 601 CGCGTTGGCTTCGCGCTGCGATCCCGACAGCTTACGCGTTCCGAGTTGCTCGCGGC 660
DB 5792 CGCGTTGGCTTCGCGCTGCGATCCCGACAGCTTACGCGTTCCGAGTTGCTCGCGGC 5851
QY 661 AAGCCCAAGGTGTTTACGATCTGTTGGAGCATTCGTTGCGCGTTCGAGGCGCATG 720
DB 5852 AAGCCCAAGGTGTTTACGATCTGTTGGAGCATTCGTTGCGCGTTCGAGGCGCATG 5911
QY 721 ATCGAAGGCAACCTCGAATCCATGCGCAATTTGGCAGCAAGCGCTCGCGCATCAGATA 780
DB 5912 ATCGAAGGCAACCTCGAATCCATGCGCAATTTGGCAGCAAGCGCTCGCGCATCAGATA 5971
QY 781 TCGATCTGGCTCGCGCGCTTACTCAAGGTCGAACCGTCCCGGATCCCGAGCTCACGCGAG 840
DB 5972 TCGATCTGGCTCGCGCGCTTACTCAAGGTCGAACCGTCCCGGATCCCGAGCTCACGCGAG 6031
QY 841 TTCGAATGGTACGTCGCGGTTCGATGAGACAGCCACTTCTTACTTCAGACGCTGGGCAA 900
DB 6032 TTCGAATGGTACGTCGCGGTTCGATGAGACAGCCACTTCTTACTTCAGACGCTGGGCAA 6091
QY 901 GTGTCGACGTCAAGAGAGCGGAGACTCTCTCGAGGAGAGATTCCACGAAATCGGTA 960
DB 6092 GTGTCGACGTCAAGAGAGCGGAGACTCTCTCGAGGAGAGATTCCACGAAATCGGTA 6151
```

```

QY 961 GGCCTCGCGCTTAACGGCTTCAATGATGACGACATCATGGACGATGATCGATGAGCGG 1020
Db 6152 GGCCTCGCGCTTAACGGCTTCAATGATGACGACATCATGGACGATGATCGATGAGCGG 6211
QY 1021 TTCTACGCTGATGATCGCGGTGGTTCGGAAGAAATCCTGTTCGAGCCGAGCCGCAATC 1080
Db 6212 TTCTACGCTGATGATCGCGGTGGTTCGGAAGAAATCCTGTTCGAGCCGAGCCGCAATC 6271
QY 1081 ATCGAGTGGCGGGGCTTGCCAGTGCAGACATCGCGGCATTCAGGAAGACGATTGA 1137
Db 6272 ATCGAGTGGCGGGGCTTGCCAGTGCAGACATCGCGGCATTCAGGAAGACGATTGA 6328

RESULT 3
AB072827
LOCUS
DEFINITION
  AB072827.1 GI:20135960
  1110 bp DNA linear BCT 20-JUN-2002
  Spingomonas sp. KAI carAa, carBa genes for putative oxygenase
  component of carbazole 1,9a-dioxygenase, putative small subunit of
  2'-aminobiphenyl 2,3-diol, partial cds.
ACCESSION
  AB072827
KEYWORDS
  Sphingomonas sp. KAI
SOURCE
  Sphingomonas sp. KAI
  Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
  Sphingomonadaceae; Sphingomonas.
REFERENCE
  1
  Haba,H., Ashikawa,Y., Saiki,Y., Yoshida,T., Nojiri,H. and Omori,T.
  Spingomonas sp. strain KAI, carrying a carbazole dioxygenase gene
  homologue, degrades chlorinated dibenzo-p-dioxins in soil
  FEMS Microbiol. Lett. 211 (1), 43-49 (2002)
  2
  22047665
  12052549
  2 (bases 1 to 1110)
  Omori,T., Haba,H. and Nojiri,H.
  Direct Submission
  Submitted (10-OCT-2001) Hiroshi Haba, The University of Tokyo,
  Biotechnology Research Center; Yayoi 1-1-1, Bunkyo-ku, Tokyo
  113-8657, Japan (E-mail:uhaba@mail.ecc.u-tokyo.ac.jp,
  Tel:81-3-5841-3070, Fax:81-3-5841-8030)
  Location/Qualifiers
    1..1110
    /organism="Sphingomonas sp. KAI"
    /mol_type="genomic DNA"
    /strain="KAI"
    /db_xref="taxon:164608"
    1..876
    /gene="carAa"
    <1..876
    /gene="carBa"
    /codon_start=1
    /transl_table=11
    /product="putative oxygenase component of carbazole
    1,9a-dioxygenase"
    /protein_id="BAB88910.1"
    /db_xref="GI:20135961"
    /translation="GDPIYHGWTYRWDGRLLVDILITNPGSVQIGRRALKTPPVEEAKG
    LIFVYVGDGEPTLIEDVPGFLDENRAIHGHRILVASNWLGAENGFDAGHFIHKN
    SILYKNDIILPLGAPGDPLQTESEVAAKPKGVYDLIGEHSPVPFEGMIEGKPAI
    HNGTSGKVAISISITLPGVLKVEPWPDPPELTQFEMVYVPVDETSHLYPQILGKVTSEK
    RGAASFREHFEHKKWGLAINGFNDDILARESMEFFYADDRGWSEELIFEPDRAIIIEW
    EGLASQHRNGIQEAR"
    876..1109
    /gene="carBa"
    876..>1109
    /gene="carBa"
    /codon_start=1
    /transl_table=11
    /product="putative small subunit of 2'-aminobiphenyl
    2,3-diol"
    /protein_id="BAB88911.1"
    /db_xref="GI:20135962"

```

```

/translation="MGTIASYPVNRMLQBELFTNPGNVLFRADREALYERYGLSSAQR
AALDEGGFGLTAVGLHPVLQMHFMLTNPAPA"

Query Match      75.8%; Score 859.2; DB 1; Length 1110;
Best Local Similarity 99.7%; Pred. No. 3.5e-186;
Matches 861; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 274 TATCAGCGCTGGACATATCGCTGGGACGATGGCGCTCGTCGATATCTCACAACCC 333
Db 13 TATCAGCGGTGGACCTATCGCTGGGACGATGGCGCTCGTCGATATCTCACAACCC 72
QY 334 GGCAGTGTGCAGATCGCGCGCGCGCTTTGAAGACGTTCCCGGTTGAAGAGGCAAGGT 393
Db 73 GGCAGTGTGCAGATCGCGCGCGCGCTTTGAAGACGTTCCCGGTTGAAGAGGCAAGGT 132
QY 394 CTTATCTTCTGTTTACCTAGGCGGACGACCAACCGCGCTTATCAAGATGTGCCGCC 453
Db 133 CTTATCTTCTGTTTACCTAGGCGGACGACCAACCGCGCTTATCAAGATGTGCCGCC 192
QY 454 GGCTTCTTGTATGAAAACCGCGCATTCACGGCCAAACATCGGCTCGTGGCTCGAACTGG 513
Db 193 GGCTTCTTGTATGAAAACCGCGCATTCACGGCCAAACATCGGCTCGTGGCTCGAACTGG 252
QY 514 CGCTTGGGTGCGAAAAACCGCTTTGATGGGGCAGCTTCTTCAATTCACAAGATTCGATC 573
Db 253 CGCTTGGGTGCGAAAAACCGCTTTGATGGGGCAGCTTCTTCAATTCACAAGATTCGATC 312
QY 574 CTGCTGAAGGGCAACGATATCATCTCGCGCTTGGCTTTGCGCTCGCGATCCCGACCAAG 633
Db 313 CTGCTGAAGGGCAACGATATCATCTCGCGCTTGGCTTTGCGCTCGCGATCCCGACCAAG 372
QY 634 CTTACCGGTTCCGAGTGTGCTGGGCAAGCCCAAGAGGTGTTTACGATCTGCTTGGCGAG 693
Db 373 CTTACCGGTTCCGAGTGTGCTGGGCAAGCCCAAGAGGTGTTTACGATCTGCTTGGCGAG 432
QY 694 CATTGCTGCGCGTGTTCGAAGGCATGATCGAAGGCAAACTGCAATTCATGGCAACATT 753
Db 433 CATTGCTGCGCGTGTTCGAAGGCATGATCGAAGGCAAACTGCAATTCATGGCAACATT 492
QY 754 GGCAGCAAGCGCTCGCCATCAGCATATCGATCTGCTGCGCGGCGCTACTCAAGGTGCA 813
Db 493 GGCAGCAAGCGCTCGCCATCAGCATATCGATCTGCTGCGCGGCGCTACTCAAGGTGCA 552
QY 814 CCGTGGCGGATCCCGAGCTCAGCAGTTCGATGTCGCGCTCGATGAGACCAAG 873
Db 553 CCGTGGCGGATCCCGAGCTCAGCAGTTCGATGTCGCGCTCGATGAGACCAAG 612
QY 874 CACCTCTACTTCCAGACGCTGGGCAAGTCTGATCTGCTGATGATCGCGGTGGTCCGAAGA 933
Db 613 CACCTCTACTTCCAGACGCTGGGCAAGTCTGATCTGCTGATGATCGCGGTGGTCCGAAGA 672
QY 934 GAGCGAGATTCCACGAAATGGGTAGGCTCGCGCTTAAAGGCTTCAATGATGACGAC 993
Db 673 GAGCGAGATTCCACGAAATGGGTAGGCTCGCGCTTAAAGGCTTCAATGATGACGAC 732
QY 994 ATCATGGCAGTGAATCGATGGAGCGCTTCTACGCTGATGATCGCGGTGGTCCGAAGA 1053
Db 733 ATCTTGGCAGTGAATCGATGGAGCGCTTCTACGCTGATGATCGCGGTGGTCCGAAGA 792
QY 1054 ATCTTGGTGGAGCGGACCGCAATCATCGAGTGGCGGGGCTTCCAGTCAGCAAT 1113
Db 793 ATCTTGGTGGAGCGGACCGCAATCATCGAGTGGCGGGGCTTCCAGTCAGCAAT 852
QY 1114 CGCGGATTTCAGGAAGCAGCTTGA 1137
Db 853 CGCGGATTTCAGGAAGCAGCTTGA 876

RESULT 4
AB095952
LOCUS
DEFINITION
  AB095952
  20017 bp DNA linear BCT 04-FEB-2003
  Janthinobacterium sp. J3 car operons, partial and complete cds.

```

```

ACCESSION      AB095952
VERSION        AB095952.1  GI:28201193
KEYWORDS
SOURCE
ORGANISM       Janthinobacterium sp. J3
               Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
               Oxalobacteraceae; Janthinobacterium.
REFERENCE
AUTHORS        Wadada, J., Nojiri, H., Nakai, S., Inoue, K., Shintani, M., Yoshida, T.,
               Habe, H. and Omori, T.
TITLE          Molecular diversity of carbazole-utilizing bacteria and their
               catabolic genes
JOURNAL
REFERENCE
AUTHORS        Inoue, K., Nojiri, H., Nakai, S., Endoh, T., Urata, M., Ashikawa, Y.,
               Saiki, Y., Yoshida, T., Habe, H. and Omori, T.
TITLE          Divergent structures of carbazole-degrading car operons isolated
               from several gram-negative bacteria
JOURNAL
REFERENCE
AUTHORS        Nojiri, H. and Inoue, K.
TITLE          Direct Submission
JOURNAL
FEATURES
source         1..20017
               /organism="Janthinobacterium sp. J3"
               /mol_type="genomic DNA"
               /strain="J3"
               /db_xref="taxon:213804"
               /country="Japan"
               /count="609"
               /gene="ORFU1"
               /gene="ORFU1"
               /gene="ORFU1"
               /function="unknown"
               /codon_start=1
               /transl_table=11
               /product="hypothetical protein"
               /protein_id="BAC56729.1"
               /db_xref="GI:28201194"
               /translations="MEHRFCTVPKVKVANKPRAYNRDCUNYARQDFPQGEAAKAL
               QIGTFRRNRKGRFTTQSPGNDKTSYPRGQYETGAKTLTIQSHCKEFGMYLT
               SAEKSKVQQAACKSIVQSAALCSIVQHCAA"
               985..1752
               /gene="ORFU2"
               /gene="ORFU2"
               /gene="ORFU2"
               /function="putative transfer protein"
               /codon_start=1
               /transl_table=11
               /product="transfer protein"
               /protein_id="BAC56730.1"
               /db_xref="GI:28201195"
               /translations="MRPFRCSMEVIEREMHTVHFVLMKGVGKMSASLLAQLI
               ERARLYCATDPTNTTFAYKALNVEHFNISDNNKVTFRFDALIIKIAFEGDCV
               VDTGASFLPMLYLNHENKIIDLESTGRKVLVHTPLVGGQWDETIRGLQSLIEFFS
               APVVWVYFGVKDGVGFKETSLYQNAKGTGLGVFLQELTAETSGODVANWAKK
               YLTFDEVLSAPETDLVTRHRLGIVKXHIPDLAKIDL"
               complement(2511..3413)
               /gene="ORFU3"
               /gene="ORFU3"
               /gene="ORFU3"
               /function="putative transposase ORFB protein"
               /codon_start=1
               /transl_table=11
               /product="transposase"
               /protein_id="BAC56731.1"
               /db_xref="GI:28201196"
               /translations="MRAFDYCHAYGVEPICVMQVAPSGVWRYAAQORNALRCAR
               VQRDDVLSVIERWQNLQVQADKVRQLRRREGTEVACTVERLNRKAGLRGVNRG

```

```

KVRTTAADTKAPCLDRNRQNAQRPNQLWSDFTYVSTWQGFVTVAVFVDFARR
IVGWRVSCSMRTDFVLDALQALYARQPERNELVHSDRGSYVSIKYSRLAAGIE
PSVSGKSDYDNALAEITNGLYKAEIITHRAPWKTRAEVLALEWVSWFNHRLLEP
LGYPPAPAEANYKQLSSQAIPA"
complement(3410..3733)
/gene="ORFU4"
complement(3410..3733)
/gene="ORFU4"
/function="putative transposase ORFA protein"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC56732.1"
/db_xref="GI:28201197"
/translations="MKQPKYSEVIERAVRLYSEAASEYSSQWAAIESIAAKIGCTP
ETLRWYRQQRDTGQRPGETTAEEERIKTLEREVRELKANEILLASAFFAQELG
RHFXP"
3924..4253
/gene="ORFU5"
3924..4253
/gene="ORFU5"
/function="putative DNA-binding protein"
/codon_start=1
/transl_table=11
/product="DNA-binding protein"
/protein_id="BAC56733.1"
/db_xref="GI:28201198"
/translations="MANIDLSGVNLAELKGLQHDIEREIKSRQHDLOKAREQILALA
QEAGSVVEELLAGGKSKTGKQVQARYQNPDSNSQVTWGRGRQPKTAELGAKGK
ALDDPRI"
4765..5106
/gene="ORFU6"
4765..5106
/gene="ORFU6"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56734.1"
/db_xref="GI:28201199"
/translations="MVGPAAWYKVMNRNRNHTIAALLVATLACNRRSPGLEGTWK
ADGLVPMITIFRPGTEAMGVIEKVDTYTNNGVVKVTKDGLMKGSAIAFTVNRNTA
NNPMYTLIR"
5647..6171
/gene="ORFU7"
5647..6171
/gene="ORFU7"
/function="putative cold shock protein"
/codon_start=1
/transl_table=11
/product="cold shock protein"
/protein_id="BAC56735.1"
/db_xref="GI:28201200"
/translations="MRLQGRVIEWNDERGFGFIQNGTSTERTFLHISALRKTGRPLL
GSLVYETIRDDKGRLOAAELGTSLRQSSGGFSASVLGLVLLIMFGYGVYVRF
NQGOSTTASVYKIVLARDALRPHPEFQCTPEKSKCSKASCAELPHQRCNVPKMD
GDGIGICEEOWCN"
6306..6818
/gene="ORFU8"
6306..6818
/gene="ORFU8"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56736.1"
/db_xref="GI:28201201"
/translations="MKIFASSLALILATVPSTAQVATPASEARPTLLRALDGSWRM
SGDVRGQVYTMGAALQCTFTMEMKDVQPAQYEVVIFGDKASQVIAHMD
NHAKYSIPHGTGQTIGNSIQLLFPYQSGNFRNTWTYHGDTSSWTFLESAQPDGSK
HFARYEVKRN"
6950..7378
/gene="ORFU9"

```

CDS
6950..7378
/gene="ORFU9"
/function="putative transposase ORFA protein"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC56737.1"
/db_xref="GI:28201202"
/translations="MARALLPEDLWLSIAHLFAHRRSPKGRPRIDDRAPTGLFV
LKTGIPWEYLPRLGGSGMTWRRLHEWMLAGWQRIHETILRLRREHDQIMWDRA
VDAASVPAPAGSHTGRNPTDRGLCKGKHLWMLMSVGFHM"
7348..7629
/gene="ORFU10"
7348..7629
/gene="ORFU10"
/function="putative transposase"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC56738.1"
/db_xref="GI:28201203"
/translations="MYDERGFPIVAOISGAOVHDSREFLPLVESVPAVKGLAGRAK
PGKLHADRAYASRAHRAWLRRRGIAARIARYGVESRERLGRWVWVERI"
7629..7861
/gene="carR"
7629..7861
/gene="carR"
/complement(7862..8617)
/function="putative transcriptional regulator of car
operon"
8617..8880
/codon_start=1
/transl_table=11
/product="regulatory protein"
/protein_id="BAC56739.1"
/db_xref="GI:28201204"
/translations="MONTNLYDQGAPKASARAVKETSBIVEEROSRQTVVEIAAKA
IKANGILFAPORVADLTVELGISRNPLRAFGRLVARGLLQVEPFGAIVARL
SRKELVDLSRVSRCLSWAARLAEINDKSPFNVYLEASIRPSTPLPNVSYLE
DNKRFSLILKLGNHSLPKLLDLQLPPIQSNAPFMYDANIYQNSTABHYELQAIL
AGDAASTRVAAHVRHTSKLLSLPDLHPFNSD"
8880..9141
/gene="ORFU12"
8880..9141
/gene="ORFU12"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56740.1"
/db_xref="GI:28201205"
/translations="MILQVEFLDNIALNAIPDCDGLLSFQAAASVCYFRKIALQNC
SNTSFYNFRYQNFLIKFLKDMYRTVVSNAHGLVKNLKG"
9141..10130
/gene

Query Match 37.2%; Score 422.6; DB 1; Length 20017;
Best Local Similarity 62.8%; Pred. No. 3.5e-86;
Matches 691; Conservative 0; Mismatches 404; Indels 6; Gaps 2;

QY 31 AGAACCAAGGTTTGGGAGCCCTTATCCGTCGAAACTCGGTTCCGAAACCATTCGAT 90
Db 10244 AGAGTAAAGGCTGGGGCCCTACGTCGATCGCAATCATTGGTAC 10303

QY 91 CCCGTTCCGCTCGGAGCGAANTCGCGGAAGGTACTCCCGTCCCGTCAAGTCTCTGGGA 150
Db 10304 CCGGTGATGTTTTCGAAGGAGATCAACGAGGGCGAGCCGAAGCGCTAAACTGCTCGT 10363

QY 151 GAGAAGATTCTGCTCAATCCGTCGGCGGCAAGCTCTATCGATCCAGACAGGTGCGCTG 210
Db 10364 GAGAACTTGCTCGTCAATCGATTCGATGGGAAGCTGATTGCTTCAAGACCGTGCCTG 10423

QY 211 CATCGCGGTGTAACGCTTTCCGACCGGGTCGAGTGTCTATTCGAAGAACACCATTCCTGC 270
Db 10424 CATCGCGGTGTCAGTTGTCTGTCAAAGTCGAGTGCAGAACGAAATCGACATCATGC 10483

AUTHORS Omori, T. and Takami, K.
TITLE OXIDASE GENE ORIGINATING FROM MICROORGANISM AND REMOVAL OF DIOXIN
JOURNAL WITH THE SAME
COMMENT Patent: JP 1998257895-A 1 29-SEP-1998;
ASAHI CHEM IND CO LTD
OS Pseudomonas sp.
PN JP 1998257895-A/1
PD 29-SEP-1998
PF 18-MAR-1997 JP 1997084401
PI OMORI TOSHIO, TAKAMI KAZUTAKA
PC C12N15/09, A62D3/00, B09C1/10, C02F3/34, C07H21/04, PC
C12N1/21//C12N9/02,
PC (C12N15/09, C12R1/38), (C12N1/21, C12R1/19), (C12N9/02, C12R1/19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key
FH Location/Qualifiers
FT source 1..1161
FT /organism='Pseudomonas sp.'
FT /strain='CA10'.
FT Location/Qualifiers
FEATURES
source 1..1161
/organism='Pseudomonas sp.'
/mol_type='genomic DNA'
/db_xref='taxon:306'
ORIGIN
Query Match 36.98; Score 419.4; DB 6; Length 1161;
Best Local Similarity 62.68; Pred. No. 2e-85;
Matches 689; Conservative 0; Mismatches 406; Indels 6; Gaps 2;
QY 31 AGAACCAAGTTTGGGACGCTTATATCCGTGCGAAACTCGGTTCCGAAACCACTTGGTAT 90
DB 37 AGAGTAAAGGCTGGGGCGCCTACGTGATCGAAGCTAGGCTTTCGCAATCATTGGTAC 96
QY 91 CCGGTCGCTCGCGGCAAGTCCGGAAGTACTCCGTTCCGCTCAAGTCTCTGGGA 150
DB 97 CCGGTGATGTTTCGAAGAGATCGAGAGGGGAGCGGAGACATAAATCTGCTCGGT 156
QY 151 GAGAAGATTCGTCAATCGCGTGGGCGCAAGTCTATGCGATCCAGGACAGTGCCTG 210
DB 157 GAGAATTCGTCAATCGTATCGATGGAAGCTGTATTGCTCAAGGACCGTGCCTG 216
QY 211 CATCGCGGTGTAAGCTTTCGACCGGTCGAGTCTATTCGAAGAACCAATATCTCTG 270
DB 217 CATCGCGGCTCAGTTGTGCGTCAAGTCTGAGTCAAAACGAAGTCAATCATGTC 276
QY 271 TGATATCACGGCTGGACATATCGTGGGACGATGGCGGCTCGTATATCTCTCAAAAC 330
DB 277 TGATACACGCTGGACCTATCGTGGGAGACGGGCTTCTGCGACATCTTGAGTAT 336
QY 331 CCGGCGAGTGTGAGATCGCGCGGCGGCTTTGAAGACGTTTCCCGTTGAAGGCGCAA 390
DB 337 CCGACAGCGCAGATCGGTGACAAAAGCTGAAAACCTTACCCAGTGCAGGAAGCAAG 396
QY 391 GGTCTTATCTTCGTTACGTAGCGAGCGGCAAGCAACCGCGCTTATCGAAGATGCGCG 450
DB 397 GGCTGCGTCTTCATTATCTTGCGGATGAGGACCTCTCTCTTGGCCCGGATACGCCA 456
QY 451 CCGGCTTCCTTGATGAAAACCGGCGCATTCAGGCGCAACATCGGTCGTGGCGCTCGAAC 510
DB 457 CCAATTTCTTGACGATGATGGAATCTCTCGGAAGAACCAATCATCAAGTCTAAC 516
QY 511 TGGCGCTTGGTGGGAAAACGCTTGTATGCGGGGACGCTCTTCATTCACAGATTGCG 570
DB 517 TGGCGCTGCTGTGGAACAGGTTTCGATCCAGGCCACATTTATTCACAAAGACTCA 576
QY 571 ATCTGCTGAAGGCGCAACGATATCATTTCTGCCGCTTGGCTTTGCGCC---TGGCGATCCC 627
DB 577 ATTCTGCTCAAGGACACGATCTTGCTTGCCACTAGGTTTGGCGCCAGGAGGGATCGA 636

QY 628 GACGAGCTTACGCGT---TCCGAGGTTGCTCGCGGCAAGCCCAAGGTGTTTACGATCTG 684
DB 637 AAGCAACAACTCGTGTGTTGACGATGACGCTCGTGGACGCAAGGGTGTTTACGATCTT 696
QY 685 CTTGGCGAGCATTCGGTGCCTGTTTGAAGGCAATGATCGAAGCAAAACCTGCAATCCAT 744
DB 697 ATTGGCGAATCGGGTCCCAAGTCTTTGAGGGAATCTATCGGGGCGGAGTGTTCGCGAA 756
QY 745 GCGAATATTGGCAGCAGCGCTCGCATACGATATCGATCTGGCTGCGGGGCTACTC 804
DB 757 GGTGCTACGGCGGAAAAAATTGTAGCGAACGATATCTCCATTTGGCTCCCGGTGTTCTC 816
QY 805 AAGTTCGAACCGTGGCGGATCCCGAGCTCAGCAGTTCGATGCTAGTCCGCTCGAT 864
DB 817 AAGTTCATCGTTCCCAATCCGACATGATGATGCTGAGTGTAGTGTGCTCCGATTCAC 876
QY 865 GAGACCAACCTCTACTTCCAGACGCTGGGCAAGTCTGACGTCAAGGAAGCGCA 924
DB 877 GAAACACACACTATTACTTCCAAACTCTTGGCAACCATGTGCCAATGACGGAACGG 936
QY 925 GACTCTTCGAGCGAGATTCACGAAATGGGTAGGCTCGGCTTAAGGCTTCAAT 984
DB 937 AAGATTTAGCAACAGAGTTTGAAGCAAGTGGAAACCGATGGCGCTCGAAGGATCAAC 996
QY 985 GATGACGACATCATGGCACGTAATCGATGAGCGCTTCTACGCTGATGATCGCGTGG 1044
DB 997 AACGATGACATCTGGGCTCGGAAAGTATGTTGATTTCTACGCCGATGATAAAGGCTGG 1056
QY 1045 TCCGAGGAATCTCTTCGAGCGGACCGGCAATCATCGATGCGGGGCTTCCAGT 1104
DB 1057 GTCACGAGATTTGTTTCGAGGTGAGGCTATCTGTCATGCGCAAGCTGCGCAGC 1116
QY 1105 CAGCACATCGGCAATTCAG 1125
DB 1117 GAACAAATCAGGTATTTCAG 1137
RESULT 6
E17293 1200 bp DNA linear PAT 28-JUL-1999
LOCUS gDNA encoding oxidase which converts dibenzo-p-dioxin to a diphenyl
DEFINITION ether triol.
ACCESSION E17293
VERSION E17293.1 GI:5711976
KEYWORDS JP 1998257895-A/2.
SOURCE Pseudomonas sp.
ORGANISM Bacteria; Proteobacteria.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Omori, T. and Takami, K.
TITLE OXIDASE GENE ORIGINATING FROM MICROORGANISM AND REMOVAL OF DIOXIN
JOURNAL WITH THE SAME
PATENT: JP 1998257895-A 2 29-SEP-1998;
ASAHI CHEM IND CO LTD
OS Pseudomonas sp.
PN JP 1998257895-A/2
PD 29-SEP-1998
PF 18-MAR-1997 JP 1997084401
PI OMORI TOSHIO, TAKAMI KAZUTAKA
PC C12N15/09, A62D3/00, B09C1/10, C02F3/34, C07H21/04, PC
C12N1/21//C12N9/02,
PC (C12N15/09, C12R1/38), (C12N1/21, C12R1/19), (C12N9/02, C12R1/19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key
FH Location/Qualifiers
FT source 1..1200
FT /organism='Pseudomonas sp.'
FT /strain='CA10'.
FT Location/Qualifiers
FEATURES
source 1..1200

/organism="Pseudomonas sp."
/mol_type="genomic DNA"
/db_xref="taxon:306"

ORIGIN

```
Query Match      36.9%; Score 419.4; DB 6; Length 1200;
Best Local Similarity 62.6%; Pred. No. 2e-85;
Matches 689; Conservative 0; Mismatches 406; Indels 6; Gaps 2;

QY 31 AGAACCAAGTTTCGGAGCCTTATATCGGTGCGAAATCGGTTCCGAAACCACTTGGTAT 90
DB 31 |||||
QY 32 AGACTAAAGCTGGCGCCCTACGTGATGCGAGCTAGCTTTCGCAATCATTTGGTAC 153
DB 32 |||||
QY 33 CCGTTCCCTCGCGAGCGAAATCGCGAAGGTAATCCCGTTCCCGTCAAGCTCCTGGGA 150
DB 33 |||||
QY 34 CCGTGTGATGTTTTCGAAGAGATCGACGAGCGGCGAGACACTAAACACTGCTCGGT 213
DB 34 |||||
QY 35 GAGAAGATTCTGCTCAATCGGTCGGCGGCAAGTCTATGCTATCCAGGACAGTGCTG 210
DB 35 |||||
QY 36 GAGACTTGTCTGCTCAATCGTATCGATGGGAGCTGTATTGCTCAAGGACCGTCCCTG 273
DB 36 |||||
QY 37 CATCGGCTGTAACGCTTTCCGACCGGCTCGAGTCTATTCCAAAGACACATATCTGC 270
DB 37 |||||
QY 38 CATCGCGCTGCCAGTTGTCGTTCAAAAGTGGAGTCAAAACGAAAGTCGACGATCACA 333
DB 38 |||||
QY 39 TGGTATCAGCGCTCGACATATCGTGGGAGATGCGCGCTCGTCAATATCCTCAAAAC 330
DB 39 |||||
QY 40 TGGTATCAGCGCTCGACATATCGTGGGAGATGCGCGCTCGTCAATATCCTCAAAAC 330
DB 40 |||||
QY 41 CCGCGAGTGTGACATCGCGCGCGCTTTGAAGACGTTCCCGTGTGAAGAGGCCAAA 390
DB 41 |||||
QY 42 CCGCAAGCGCACAGATCGGTGCGCAAAAGCTGAAACTTACCAGTGCAGGAGCCCAAG 453
DB 42 |||||
QY 43 GGTCTTATCTCTGTTACGTAGCGGAGCGGCAACCAACGCGCTTATCGAAGTGTGCG 450
DB 43 |||||
QY 44 GGTGCTGCTTCAATTTATCTGCGGATGCGGACCTCTCTCCCTTGGCCCGCATACCCA 513
DB 44 |||||
QY 45 CCCGCTCTCTTGTGATAAAGCGCGCATTCACGCGCAACATCGCTCGTGGCTCGAAC 510
DB 45 |||||
QY 46 CCCAATTTCTTGTGACGATGACATGAAATCTCTCGGGAAGCAACAAATCATCAAGTCTAAC 573
DB 46 |||||
QY 47 TGGGCTTGGTGGGGAAGCGCTTTGATCGCGGGGACGCTTCTTCAATCAAGAAATCG 570
DB 47 |||||
QY 48 TGGGCTTGGTGGGGAAGCGCTTTGATCGCGGGGACGCTTCTTCAATCAAGAAATCG 570
DB 48 |||||
QY 49 ATCTGTTGAAAGGCAACGATATCATCTGCGCTTGGCTTTGGCGC---TGGCGATCCC 627
DB 49 |||||
QY 50 ATCTGTTGAAAGGCAACGATATCATCTGCGCTTGGCTTTGGCGC---TGGCGATCCC 627
DB 50 |||||
QY 51 GACGAGTTACGCT---TCCGAGTTGCTCGCGGCAAGCCCAAGGTGTTTACGATCTG 684
DB 51 |||||
QY 52 AAGCAACAACTCGTGTGTTGACGATGACGTCGTCGAGCGCAAGGGTGTTTACGATCTT 753
DB 52 |||||
QY 53 CTTCGAGCATTCGGTGGCGGTTTTCGAGGCGATGATCGAAGCAACCTCGCAATCCAT 744
DB 53 |||||
QY 54 ATTGGCAACATGGGTTCCAGTGTGTTGAGGAACTATCGGGGCGAAGTGTGTCGCGAA 813
DB 54 |||||
QY 55 GGCAACATTCGACAAAGCGCTGCCCATCAGCATATCGATCTGCTGCTCGCGGCGTACTC 804
DB 55 |||||
QY 56 GGTGCTACGCGCAAAATTTGATGCGCAACGATATCTCCATTTGCTCCCGGGTGTCTC 873
DB 56 |||||
QY 57 AAGTTCGAACCGTGGCGGATCCCGAGCTCAGCAGTTCGATGATGATGATGATGATGAT 864
DB 57 |||||
QY 58 AAGTTCGAACCGTGGCGGATCCCGAGCTCAGCAGTTCGATGATGATGATGATGATGAT 864
DB 58 |||||
QY 59 GAGACCAACCACTTACTTCCAACTCTTGGCAACCATGTGCAATGATGATGATGATGAT 924
DB 59 |||||
QY 60 GAGACCAACCACTTACTTCCAACTCTTGGCAACCATGTGCAATGATGATGATGATGAT 924
DB 60 |||||
QY 61 GACTCTTCGACGAGAAATTCAGAAATATGGGTAGGCTCGGCTTAAACGCTTCAAT 984
DB 61 |||||
QY 62 AAGAATTACGAACGAAGTTCGAAAGCAAGTGGAAACCGATGGCGCTCGAAGGATTCAC 1053
DB 62 |||||
```

```
QY 985 GATGACGACATCATGCGCACGTGAATCGATGGAGCCGTTCTACGCTGATGATCGCGTTGG 1044
DB 1054 AACGATGACATCTGGCTCGGAGCTATGGTGGATTTCTACGCCGATGATAAGGCTGG 1113
QY 1045 TCCGAGAAATCTCTTTCGAGCCGACCGCGCAATCATCGAGTGGCGGGGCTTCCAGT 1104
DB 1114 GTCAACGAGATTTTTCGAGGTGACGAGGCTATCGTGGCATGCGCGCAAGCTGCGGAGC 1173
QY 1105 CAGCAATTCGCGCATTCAG 1125
DB 1174 GACCAATCAGGCTATTACG 1194

RESULT 7
AB001723 6880 bp DNA linear BCT 07-JUL-1998
LOCUS Pseudomonas stutzeri carbazole catabolic genes, complete cds,
clone:PSB185.
AB001723
AB001723.1 GI:3293057
carbazole dioxygenase subunit; ferredoxin reductase; unknown;
ferredoxin; 2-hydroxy-6-oxo-6-(2'-aminophenyl)hexa-2,4-dienoate
hydrolase; 2'-aminobiphenyl-2,3-diol 1,2-dioxygenase.
Pseudomonas stutzeri
Pseudomonas stutzeri
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (sites)
REFERENCE
Ouchiya.N., Miyachi,S. and Omori,T.
Cloning and nucleotide sequence of carbazole catabolic genes from
Pseudomonas stutzeri strain OM1, isolated from activated sludge
J. Gen. Appl. Microbiol. 44, 57-63 (1998)
2 (bases 1 to 6880)
Ouchiya.N.
Direct Submission
Submitted (10-MAR-1997) Naoki Ouchiya, Chemicals Inspection &
Testing Institute, Kurume Research Laboratories; Chuo-machi 19-14,
Kurume, Fukuoka 830, Japan (E-mail:KYS04304@niftyserve.or.jp,
Tel:0942-34-1500, Fax:0942-39-6804)
Location/Qualifiers
1..6880
/organism="Pseudomonas stutzeri"
/mol_type="genomic DNA"
/strain="OM1"
/db_xref="taxon:316"
/clone="psb185"
571..1725
/codon_start=1
/transl_table=11
/product="carbazole dioxygenase subunit"
/protein_id="BAA31266.1"
/db_xref="GI:3293058"
/translation="MANVDEAILKRVKGMWAPYVDKALGRNHWYPMVFSKEIDGEPK
TLKLGLENLVNRIDGKLYCLKRCHLHGVOLSVKVECKTKSTITCWYHMYRWEDG
VLCDILNPTSAIQGRQKLYPVOEAKGVFIYLGDDGPPPLARDTTPNPLDDMEI
LGNKQIISKNRWLAVENTGDFSHYIHKDSILVKNDLALPLGFAFGDRKQOTRVD
DVGKRGYDILGHEGVFVFGTIGGVRREGAYGEKIVANDISILWPGVLKNPFP
NPMNQFWYVIDENTHYFQTLGRCANDEERNKYQEFESKWKWALGFGNNDI
WAREAMVDFYADDKGMVNEILFEVDEAIVAWRKLASEHNGQITQAHVSG"
1837..2991
/codon_start=1
/transl_table=11
/product="carbazole dioxygenase subunit"
/protein_id="BAA31267.1"
/db_xref="GI:3293059"
/translation="MANVDEAILKRVKGMWAPYVDKALGRNHWYPMVFSKEIDGEPK
TLKLGLENLVNRIDGKLYCLKRCHLHGVOLSVKVECKTKSTITCWYHMYRWEDG
VLCDILNPTSAIQGRQKLYPVOEAKGVFIYLGDDGPPPLARDTTPNPLDDMEI
LGNKQIISKNRWLAVENTGDFSHYIHKDSILVKNDLALPLGFAFGDRKQOTRVD
DVGKRGYDILGHEGVFVFGTIGGVRREGAYGEKIVANDISILWPGVLKNPFP
NPMNQFWYVIDENTHYFQTLGRCANDEERNKYQEFESKWKWALGFGNNDI
WAREAMVDFYADDKGMVNEILFEVDEAIVAWRKLASEHNGQITQAHVSG"
3049..3321
/codon_start=1
/transl_table=11
/product="carbazole dioxygenase subunit"
/protein_id="BAA31267.1"
/db_xref="GI:3293059"
/translation="MANVDEAILKRVKGMWAPYVDKALGRNHWYPMVFSKEIDGEPK
TLKLGLENLVNRIDGKLYCLKRCHLHGVOLSVKVECKTKSTITCWYHMYRWEDG
VLCDILNPTSAIQGRQKLYPVOEAKGVFIYLGDDGPPPLARDTTPNPLDDMEI
LGNKQIISKNRWLAVENTGDFSHYIHKDSILVKNDLALPLGFAFGDRKQOTRVD
DVGKRGYDILGHEGVFVFGTIGGVRREGAYGEKIVANDISILWPGVLKNPFP
NPMNQFWYVIDENTHYFQTLGRCANDEERNKYQEFESKWKWALGFGNNDI
WAREAMVDFYADDKGMVNEILFEVDEAIVAWRKLASEHNGQITQAHVSG"
3049..3321
```

```
/codon_start=1
/transl_table=11
/product="2'-aminobiphenyl-2,3-diol 1,2-dioxygenase"
/protein_id="BAA31268.1"
/db_xref="GI:3293040"
/translations="MARYBVDRLIQMSKKEGLIGRVIDTPSDVFEYGLTTPPERTAL
LEGPALASIGVHPILQMHLYMKNPENWATHSVIKOYSDMLKGA"
3321..4130
/codon_start=1
/transl_table=11
/product="2'-aminobiphenyl-2,3-diol 1,2-dioxygenase"
/protein_id="BAA31269.1"
/db_xref="GI:3293061"
/translations="MGIKVAAGTSHILMSPKGCESAAVVNGIAELGRRLKEARPD
VLIITSDMFNINLQMSKKEGLIGRVIDTPSDVFEYGLTTPPERTAL
DGLDLCQAEYSDHIGIMIPILFMKKEIPVVPVIVINTDIPISARRCVLAESIRQ
AIEKRPDGRVAVAGGLSHWLCVPRHGEVSEKFDHVMDELARGNAEELVAMGNE
AIIIDGNGAGVEILTIVMAVAASEASGEKVFYEATQWTFGIGGMEPHVK"
4120..4992
/codon_start=1
/transl_table=11
/product="2'-aminobiphenyl-2,3-diol 1,2-dioxygenase"
/protein_id="BAA31270.1"
/db_xref="GI:3293062"
/translations="MINKAEQISEKSESAVVERFVNAGVETRYLEAGKGPVILLHG
GGAGSEGNWRVVIPIARHVRVIAIMDLGPKAKPDIETQDRIIRHLHDFIKAM
NFGKYSIVGNSMGATGLGVSLHSELVNLVLMGSLVVEIHEHEDURPLIYNDFTR
EGMVLKLTNDGKIDIDAMINSRYIATDEATKAYVATWMTREOGGLYDFEPI
RKVPFPLVHVHDKVVPVETAYFELDLIDDSWGYIIPHCGNAMIHEPDPANATL
SPLSRADITRAAA"
5013..5336
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="BAA31271.1"
/db_xref="GI:3293063"
/translations="NNQIWLKVCASDMQFGTIRRVNRVGAAPLAVRVGDOFYATED
TCTHGIASLSEGLDGVIECPFHGAVNCTGMPASSPCTVPLGVFEVKEGEVYV
AVEBK"
5370..5687
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA31272.1"
/db_xref="GI:3293064"
/translations="MADLSVITERVTKAVGENSLDVAVKDFEPBGVTHIDGMSIPN
RVSNELPSDITIKILENFEKILNQDLQPKALATGRVRLGDIRIATRLDKVFLGA
PSM"
5717..6706
/codon_start=1
/transl_table=11
/product="ferredoxin reductase"
/protein_id="BAA31273.1"
/db_xref="GI:3293065"
/translations="MYQLIEQAPGTCGSKSLVLSALANGIPIFYECASGGCGVCK
FELLEGVSMWPDAGLSRREKRNHLACQVALSDIRIKVAVQDKYVETIPIIR
MEAEVVEVRLTHLLSVLRDTPANFLPGFCILVEAEPLGVVRAVYANLKNPEG
INFEYIKRYPGTSPFLPNRKEGARLEFGPMGTSPFRPGTKRSLCIGGAGLSY
AAAIARASRETDKPKLFGSRTPRDAVRWIDIDIDEDKLEVOAVMEDTSLWQGP
TGFHQVDAALLETLPVEYIYLAGPPVVDATVRMLLKGVPRIQHFDAPF"
```

ORIGIN

```
Query Match 36.9%; Score 419.4; DB 1; Length 6880;
Best Local Similarity 62.6%; Pred. No. 1.9e-85;
Matches 689; Conservative 0; Mismatches 406; Indels 6; Gaps 2;

QY 31 AGAACCAAGGTTGGGAGCCTTATATCCGTGCGAACTCGGGTTCGGAACCATTCGTAT 90
Db 601 AGAGTAAAGGCTGGGCGCCCTACGTGGATCGAGAGCTAGGCTTTCGCAATCATTTG 660
QY 91 CCGCTTCGCTCGGAGCGGAAATCGCGGAAAGGTAATCCCGTTCCTCGCTCAAGTCTCC 150
```

RESULT 8
DB9064

LOCUS	D89064	6881 bp	DNA	linear	BCT 12-SEP-1997				
DEFINITION	Pseudomonas sp. gene for carbazole 1,9a-dioxygenase, meta-cleavage enzyme, meta-cleavage compound hydrolase, complete cds.								
ACCESSION	D89064								
VERSION	D89064.1	GI:2317677							
KEYWORDS	reductase component of carbazole 1,9a-dioxygenase; ferredoxin component of carbazole 1,9a-dioxygenase; meta-cleavage compound hydrolase; meta-cleavage enzyme; terminal dioxygenase component of carbazole 1,9a-dioxygenase.								
SOURCE	Pseudomonas sp.								
ORGANISM	Pseudomonas sp.								
REFERENCE	Bacteria; Proteobacteria.								
AUTHORS	Sato,S.I., Ouchiya,N., Kimura,T., Nojiri,H., Yamane,H. and Omori,T.								
TITLE	Cloning of genes involved in carbazole degradation of Pseudomonas sp. strain CA10: nucleotide sequences of genes and characterization of meta-cleavage enzymes and hydrolase								
JOURNAL	J. Bacteriol.	179 (15),	4841-4849	(1997)					
MEDLINE	97386424								
PUBMED	9244273								
REFERENCE	2 (sites)								
AUTHORS	Sato,S.I., Nam,J.W., Kasuga,K., Nojiri,H., Yamane,H. and Omori,T.								
TITLE	1,9a-dioxygenase in Pseudomonas sp. strain CA10								
JOURNAL	J. Bacteriol.	179 (15),	4850-4858	(1997)					
MEDLINE	97386425								
PUBMED	9244274								
REFERENCE	3 (bases 1 to 6881)								
AUTHORS	Omori,T.								
TITLE	Direct Submission								
JOURNAL	Submitted (14-NOV-1996) Toshio Omori, The University of Tokyo, Biotechnology Research Center; Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan (E-mail:aseigv@hongo.ecc.u-tokyo.ac.jp, Tel.03-3812-2111(ex.3067), Fax:03-5802-3326)								
FEATURES	Location/Qualifiers								
source	1..6881								
	/organism="Pseudomonas sp."								
	/mol_type="genomic DNA"								
	/strain="CA10"								
	/db_xref="taxon:306"								
gene	572..1726								
	/gene="carAa"								
CDS	572..1726								
	/gene="carAa"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21728.1"								
	/db_xref="GI:2317678"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21729.1"								
	/db_xref="GI:2317679"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21730.1"								
	/db_xref="GI:2317680"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21731.1"								
	/db_xref="GI:2317681"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21732.1"								
	/db_xref="GI:2317682"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21733.1"								
	/db_xref="GI:2317683"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21734.1"								
	/db_xref="GI:2317684"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21735.1"								
	/db_xref="GI:2317685"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21736.1"								
	/db_xref="GI:2317686"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21737.1"								
	/db_xref="GI:2317687"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21738.1"								
	/db_xref="GI:2317688"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21739.1"								
	/db_xref="GI:2317689"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21740.1"								
	/db_xref="GI:2317690"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21741.1"								
	/db_xref="GI:2317691"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21742.1"								
	/db_xref="GI:2317692"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21743.1"								
	/db_xref="GI:2317693"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21744.1"								
	/db_xref="GI:2317694"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21745.1"								
	/db_xref="GI:2317695"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21746.1"								
	/db_xref="GI:2317696"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21747.1"								
	/db_xref="GI:2317697"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21748.1"								
	/db_xref="GI:2317698"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21749.1"								
	/db_xref="GI:2317699"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	1838..2392								
	/gene="carAb"								
	/codon_start=1								
	/transl_table=11								
	/product="terminal dioxygenase component of carbazole 1,9a-dioxygenase"								
	/protein_id="BAA21750.1"								
	/db_xref="GI:2317700"								
	/translation="MANYDEAILKRVKGMVYDPAKLGFRNHWYFVMPFSKEIDGEPK TLKLLGENLLVNRIDGKLYCLKRCLHRGVLSVKVECKTKSTITICWHAWYRWEDG VLCDLTNPTAQIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPPNFLDDMEI LGKNQI IKSNWLA VENGDPDSHIYHKDSILVKNQNDLALPLGAPGDRKQQRVVD DDVGRKGVYDLIGHGVPVEGTIGGVREGAYGEKIVANDISILWPGVLKVNPPF NPDMMQFWEYFIDENTHYFOTLGKPCANDEERKNYQEFESKWKPMALGEGFNDDI WAREAMVDYADDKGWNNEILFEVDEAIVAWRKLASEHNQGIQTQAHVSG"								
	1838..2392								
gene	1838..2392								
	/gene="carAb"								
CDS	183								

Qy 211 CATCGCGGTGTAAGCTTTTCGACCGGGTCGAGTGTCTATTCCAAAGAACCATATCTCTGC 270
|||
Db 782 CATCGCGCGTCCAGTGTCTCGGTCAAAAGTCGAGTGCAGAAACGAAAGTCGAGTACATGTC 841
|||
Qy 271 TGGTATCACGCGTGGACATATCGTGGAGCGATGGCGCTCTGCTGATATCTCCACAAAC 330
|||
Db 842 TGGTACCAACGCTGGACATATCGTGGAGAGAGCGGCTCTGTCGACATCTTTCGAGAT 901
|||
Qy 331 CCGCGGAGTGTGAGATCGCGCGCGCTTTGAAGAGTGTCCCGTGAAGAGGCCAA 390
|||
Db 902 CCGAAGCGCACAGATCGGTGCAAAAGAGTGAAGATTTACCCAGTGCAGGAAGCCAG 961
|||
Qy 391 GGTCTTATCTGTTTACGTAGGCGAGCGGACCAACCGGCTTATCGAGATGTCCG 450
|||
Db 962 GGTGCTCTTATCTTATCTTGGGATGGGACCTCTCTCTTGGCCCGCATACGCCA 1021
|||
Qy 451 CCGCGCTTCTTGATGATAAAGCGCGCCATTCACGCGCAACATCGGCTCGTGGCTCGAAC 510
|||
Db 1022 CCAATTTCTTGACGATGATGAAATCTCTGGGAAGAACCAATCATCAAGTCTAAC 1081
|||
Qy 511 TGGCGCTTGGTGGGAACCGCTTTGATGCGGGGACGCTCTTCAITCACAGATTCG 570
|||
Db 1082 TGGCGCTCTGCTGTGGAACCGGTTTCGATCCGAGCCACATTTATATTCRAAGACTCA 1141
|||
Qy 571 ATCTGTGTGAAGGCAACGATATCATTTCTCGCGCTTGGCTTTCGCGC---TGGCGATCCC 627
|||
Db 1142 ATTCTGTCAAGGACACGATCTTCCCTGCCACTAGTTTCGCGCCAGAGGGATCGA 1201
|||
Qy 628 GACGAGCTTAGCGT---TCGAGGTTGCTGCGGGCAAGCCAAAGTGTTTACGATCTG 684
|||
Db 1202 AAGCAACAAACTCTGTGTGTGACGATGACGTCGTCGGAGCGAAGGTGTTTACGATCTT 1261
|||
Qy 685 CTTGCGGAGCATTCGGTGCGCGTTCGGAAGGATGATCGAAGCAACCTGCAATCCAT 744
|||
Db 1262 ATTGCGGACATGGGTCCCACTGTTTGAGGAGACTATCGGGGCGAAGTGTCGCGGAA 1321
|||
Qy 745 GGCACATTTGCGACGACGCGTCCCATACGATATCGATCTGGCTGCGCGGCGTACTC 804
|||
Db 1322 GGTGCTTACGCGCAAAATTTGTAGCAACGATATCTCCATTTGGCTCCCGGGTGTCTC 1381
|||
Qy 805 AAGTGTGAACGCTGCGGATCCCGAGTCCAGCTACGAGTTCGAATGATGTCGCGGTCGAT 864
|||
Db 1382 AAGTCAATCCGTTCCCAATCCGCAATGATGATGATGATGATGATGATGATGATGATG 1441
|||
Qy 865 GAGACCGACCACTTACTTCCAGACGCTGGGCAAGTCTGATCGTCAAAAGGAGCGCA 924
|||
Db 1442 GAAACACACACTTACTTCCAACTCTTGGCAACCATGTGCCAATGACGAGGAACGG 1501
|||
Qy 925 GACTCTTCGAGCGAGATTCACGAAATGGGTAGGCTTCGCGCTTACGCTTCAAT 984
|||
Db 1502 AAGAAATTACGAACAGAGTTTGAAGCAAGTGGAAACCGATGGCGCTCGAAGGATCAAC 1561
|||
Qy 985 GATGACGACATCATGGCAGCTGAATCGATGGAGCGGCTTCTACGCTGATGATCGCGTTGG 1044
|||
Db 1562 AACGATGACATCTGGCTCGGAGCTATGCTGGATTCTACGCCGATGATAAAGGCTGG 1621
|||
Qy 1045 TCCGAGAAATCTGTTTCGACCGGACCGCGCATCATCGAGTGGCGGGGCTTCCAGT 1104
|||
Db 1622 GTCACGAGATTTGTTTCGAGGTGACGAGGCTATCGTGGCATGGCCAAAGCTGGCGAGC 1681
|||
Qy 1105 CAGCACATTCGCGGCAATTCAG 1125
|||
Db 1682 GAACACATCAGGGTATTCAG 1702
|||

RESULT 9
AB047548
LOCUS
DEFINITION
Pseudomonas resinovorans strain CA10 DNA, linear BCT 03-MAY-2002
region.
ACCESSION
AB047548
VERSION
AB047548.1 GI:3094152

KEYWORDS

SOURCE
ORGANISM

Pseudomonas resinovorans
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE

AUTHORS
Sato, S.I., Ouchi, Yama, N., Kimura, T., Nojiri, H., Yamane, H. and
Omori, T.

TITLE

Cloning of genes involved in carbazole degradation of Pseudomonas
sp. strain CA10: nucleotide sequences of genes encoding carbazole
of meta-cleavage enzymes and hydrolase
J. Bacteriol. 179 (15), 4841-4849 (1997)

JOURNAL

MEDLINE
97386424

REFERENCE

AUTHORS
Sato, S.I., Nam, J.W., Kasuga, K., Nojiri, H., Yamane, H. and Omori, T.

Identification and characterization of genes encoding carbazole
1,9a-dioxygenase in Pseudomonas sp. strain CA10
J. Bacteriol. 179 (15), 4850-4858 (1997)

JOURNAL

MEDLINE
97386425

REFERENCE

AUTHORS
Nojiri, H., Sekiguchi, H., Maeda, K., Urata, M., Nakai, S., Yoshida, T.,
Habe, H. and Omori, T.

Genetic characterization and evolutionary implications of a car
gene cluster in the carbazole degrader Pseudomonas sp. strain CA10
J. Bacteriol. 183 (12), 3663-3679 (2001)

JOURNAL

MEDLINE
21264379

REFERENCE

AUTHORS
Nojiri, H., Omori, T. and Habe, H.

Direct Submission
Submitted (21-AUG-2000) Hideaki Nojiri, The University of Tokyo,
Biotechnology Research Center, 1-1-1 Yayoi, Bunkyo-ku, Tokyo

113-8657, Japan (E-mail: anojiri@mail.ecc.u-tokyo.ac.jp,
Tel: 81-3-5841-3064, Fax: 81-3-5841-8030)

FEATURES

source

1. .44286
/organism="Pseudomonas resinovorans"
/mol_type="genomic DNA"
/strain="CA10"
/db_xref="taxon:53412"
complement(214. .1929)
/gene="ORF31"
complement(214. .1929)
/function="putative membrane spanning protein"
/codon_start=1
/transl_table=11
/product="putative ABC transporter subunit"
/protein_id="BAB32742.1"
/db_xref="GI:3094153"

gene

CDS

/translation="MGVINLAHGFYVWGAFRCPTPRERLVLRVGGSGSARPYTR
CLSKPACAAATTAITSTRYMTLAVFTNELITVLFGRSPFAMPTPEWYSSFVEVL
PLMTFVSRLAIFIAAGAAVAIAMLLINHLRAGADDEHMGALGVNIARLYT
LVFVFGSVLGLAGPAAAPLLSVEIGMEKVLITTFVIVVGVGSRGALAGALLIG
MVDGIRAVIPOLLQLPAEVSGLSGELISAGAVIMAVVLVVRPGLPARALRR
TPCSRVIDLLCLATFALLPLASLHPEPLVSLFTLALYGMASLIDLIGYGM
ISFGRAAFGLGIVVGVGFTSQAFLWGDGNSALLVWPLSLLYCALFAVVG
LSLRTSGVQFIMTLAFQMLFYVLSLSLGGDDGILLISRLTDFGIDLPVQFY
LCGLLVAWLFCRLRVLRFGYLRGKQSRERISILGLHPLRGLGAFVAGVGGG
LAGILWANYAMFVSPDMGAMHKSGLMAMVILGGVGTLLGLPVLGAHVILGLEQCSACG
PSTGC"

gene

CDS

complement(2085. .3245)
/gene="ORF30"
complement(2085. .3245)
/function="putative ABC transporter substrate binding
protein"
/codon_start=1
/transl_table=11
/product="putative ABC transporter subunit"
/protein_id="BAB32743.1"

```

/db_xref="GI:13094154"
/translation="MKQLVGTLLAPVAVPAPRALPRAQDPVVRIGFTITLSTPAGVLGED
SRDQLAIDDEGGKGGVPSLVVDDGLQLAKAKQIIDRMSLDGSLYGTGVIFSNV
LAAVNTAKDGYISNTPSPSSLAGCKPKFVFSVQNDTIHEMAGVAAANDLQYKK
VILAPNTQGRDAGEGRKTFKGEVTEIYTKLQDLDFSELARVSLPADAIQFPH
GGAINFAKOYSSGLNKSIPMVVPVFSMDERMLAAGDVAKGNNIVTLWNPSTNPA
NOAFKAFNDKYKVPYVYAAQTVDTABELIGALKTVDGNLKDPEALASALRNVTFDS
IRGDPAKGNHADVILLRVADANGKLVQVPQTIIAKAQVFAAEACAL"
gene
CDS
complement(3405..4385)
/gene="tnpA3"
/gene="tnpA3"
/gene="tnpA3"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB32744.1"
/db_xref="GI:13094155"
/translation="MKQMTFADAAYAGKQKTRKELFLIEMDRVVPWKGLIALIEPHY
PKGEGRPAYPLMAMLRVLLQNMFGYSDPAMEALYETTILROFAGLNLRIEDETT
ILNFRLLKHELALGILAVINGYLGRGLSLRQGTIVDATLINAPSTKNKQGRDP
EMHQTKGNQYIFGMAKHIGADDESLVHSVGTAAANVADVQVQDKLHGDENMVGAD
AGYTVGERKPEHEGRVIVQIAARRSTYKLDKRSALYKAKKIEKAKQVRAKVEHS
FRVIRKQGYTKRFRGLAKNTAQLVTILFALSNLMMARRHLLTNAGEVRL"
gene
CDS
complement(4543..5571)
/gene="antC"
/gene="antC"
/gene="antC"
/codon_start=1
/transl_table=11
/product="reductase component of anthranilate
1,2-dioxygenase"
/protein_id="BAB32745.1"
/db_xref="GI:13094156"
/translation="MSHKVAFSFDAGKTLFPFVQSHETLLDAALRNGINIPLDCREGV
GTCQGRCEGVSQYDDYDEALSAEDLRQKMLTCQTRVSSDAFVDFPSSLCNAP
GFLKGTQVAVQSPNTAILHLHDAGADQDLPFGQYARLQIPGTHGASSISFAN
RPSNGNQFTIRLLPDGLMGNVIRERSLIGDEILLEAPGAFYLRHVDKPLVPAVG
KPLSLDLEIAERGGCGHPVHLVYGVRAADLCETORIAAVERIPGPRLPFGSS
ATPARTWKGKGYITEHFDLAEFRDQALDMLVLCGPPPWVESIKQWLDQTALDQRLYY
BKFTQSN"
gene
CDS
complement(5590..6081)
/gene="antB"
/gene="antB"
/gene="antB"
/codon_start=1
/transl_table=11
/product="small subunit of terminal oxygenase component of
anthranilate 1,2-dioxygenase"
/protein_id="BAB32746.1"
/db_xref="GI:13094157"
/translation="MNSOLQVRVEQFLYRNABCAQDQWDVTLVSMFSDSDSEHFLPQWD
SEHEVTRDPKRAMSLIYVNRGGLEDVFRITGKSASTVPMPTLHLSNVRIPALA
DELELVQVNWHTLYRLTQAQPFGRATYRLRPGESWKIARKVLLINDTINSVLDF
YHL"
gene
CDS
complement(6084..7487)
/gene="antA"
/gene="antA"
/gene="antA"
/codon_start=1
/transl_table=11
/product="large subunit of terminal oxygenase component of
anthranilate 1,2-dioxygenase"
/protein_id="BAB32747.1"
/db_xref="GI:13094158"
/translation="MSSARSVEQWKNFIEGCLDFRPADGVFRIARDIFTEPDLFQLEM
ELIFKNWIVACHSEIAKNDPMTIRAGROPMTIIRDGESQLNALINACORHRTITL
RVGNQSTFTICFPFANCYKSDGRVLKVAQPEPFGDKATRGKAKRIEISYGVTF
ISLDVGDQDSQBFQSDAKVFFDMMVAQSPTEGEVLEVPKSAITYDGNWKLQNGENGLD
GYHVSIVHYNVATVQHRQVNSSEKGLTSADTLDYSLKLAGDKETDDGWFAPFHNGSV

```

```

LPSMDPNPVRPOYATIMPRLLQEQYQGGKQAKQMMHRLNLTNPSPMFFLDQISSQURI
IREIANKTEINSFCILGVVGSADRENIROFDFNFYSGMGTDPDLVFEPRQKGF
QARLEWSDISRGQVKGWVBGTANSETIGTSPVLITGTEFTHEGLYVNHQGNWQRFLE
GLARKAQOPLKLRV"
complement(7829..8809)
/gene="tnpA2"
complement(7829..8809)
/gene="tnpA2"
/notes="tnpA2"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB32748.1"
/db_xref="GI:13094159"
/translation="MKQMTFADAAYAGKQKTRKELFLIEMDRVVPWKGLIALIEPHY
PKGEGRPAYPLMAMLRVLLQNMFGYSDPAMEALYETTILROFAGLNLRIEDETT
ILNFRLLKHELALGILAVINGYLGRGLSLRQGTIVDATLINAPSTKNKQGRDP
EMHQTKGNQYIFGMAKHIGADDESLVHSVGTAAANVADVQVQDKLHGDENMVGAD
AGYTVGERKPEHEGRVIVQIAARRSTYKLDKRSALYKAKKIEKAKQVRAKVEHP
FRVIRKQGYVQVFRGLAKNTAQLVTILFALSNLMMARRHLLTNAGEVRL"
9062..10318
/gene="ORF24"
9062..10318
/gene="ORF24"
/function="channel-forming protein"
/codon_start=1
/transl_table=11
/product="PhaK-like protein"
/protein_id="BAB32749.1"
/db_xref="GI:13094160"
/translation="MFQASWFRGMQTSALALMTPLAQADFVEDSQLALGRNFYI
DRFDKGNPNKSRVSGTQGFDLRPNVSGYTEGLQFALDASQYAVRLDGGGGRGPD
IIPYSTRORQATDYGRGYTAKRVYSKTELXGHEHPTLPVAYTDDSQLVTVYHGT
LIESKEVDKLTLTGSGFTSISGBSSNSKMLXGPDVQRESDDGLNFGSAYFTPA
LGITYYQGLEDIYRCHYLGATYLDLGSYALKTDLRYFNKKEGDDALYICIDNRSY
LMTURKKAQSGFLGYQRMGLDSTPTTLNGYAPQYPLVNVSAVAFVKNKESWQARY
DYDFAGLGLPGLRLMTRYMRGSDIDRGALSDNSESRNTYLSVYVQEGPLKGLVLR
NINVRHNGHNDENRLITTYTWKFW"
10639..11655
/gene="ORF23"
10639..11655
/gene="ORF23"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="BAB32750.1"
/db_xref="GI:13094161"
/translation="NMSTSLDVHVRDIRIERYDLNACARQWAGICGPHLLKTOYPNRI
QFHHSANVLKSMSTILGVIEYDVTGVIEDVEHLNYSLSPLVGEQHLKDKGCOVT
Query Match 36.9%; Score 419.4; DB 1; Length 44266;
Best Local Similarity 62.6%; Pred. No. 1.9e-85;
Matches 689; Conservative 0; Mismatches 406; Indels 6; Gaps 2;
Qy 31 AGAACCAAGTTTGGGAGCCTTATATCCGTCGCAAACTCGGTTCCGAAACCATTCGTAT 90
Db 28539 AGATAAAGCTGGGCGCCCTACGTGATGAGAGCTAGGCTTCGCAATCATTCGTAC 28598
Qy 91 CCGCTTCGCCTCGCGAGCGAAATCGCGAGGCTACTCCCGTTCCCGTCAAGCTTCCTGGGA 150
Db 28599 CCGGTGATGTTTTCGAAAGAGATCGACGAGGCGGACCGAAGACACTAAACTGCTCGGT 28658
Qy 151 GAGAGAGTTCTGCTCAATCCGCTGGGCGCAAGTCTATCGATCCAGGACAGTCCCTG 210
Db 28659 GAGAACTTGTCTGCTCAATCGATCGATGGGAAGCTGATTCGCTCAAGGACCGCTCGCTG 28718
Qy 211 CATTCGGGTGTAACGGCTTTCGACCGCGGTGAGTGTCTATTCAGAAACACCATATCCTGC 270
Db 28719 CATCGCGGCTCCAGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 28778
Qy 271 TGGTATCACGGCTGGACATATCGCTGGGACGATGCGCGGCTCGCTGATATCTCACAAC 330
Db 28779 TGGTATCACCGCTGGACCTATCGCTGGGAGACGCGGCTTCTGTGCGACATCTTTGACGAAT 28838

```

```
QY 331 CCCGCGAGTGTGCAGATCGCGCGCGCGCTTTGAAGACGTTCCCGGTTGAAGAGCGCCAAA 390
D 28839 CCGACAGCGCACAGATCGGTGCGCAAAAGCTGAAACTTACCAGTGCAGGAAGCCAA 28898
QY 391 GGTCTTATCTTCGTTTACGTAGGCGGCGGCGGCAACCAAGCGCGCTTATCGAAGATGTCGG 450
D 28999 GGCTGCGTCTTCACTTATCTTGGCGATGCGGACCTCTCTCCCTTGGCCCGGATACGCCA 28958
QY 451 CCCGCGCTTCCCTTGATGAAACCGGCGCCATTTCAGCGCCAAATCGCGCTCGTGGCCTCGAAC 510
D 28959 CCCAATTTCCCTTGACGTACATGAAATCTCTCGGAGAACCAATCATCAAGTCTAAC 29018
QY 511 TGGCGCTTGGGTGCGGAAACCGCTTGAATCGGCGGCGACGCTTCTCATTCACAGAATCG 570
D 29019 TGGCGCTTGGGTGCGGAAACCGCTTGAATCGGCGGCGACGCTTCTCATTCACAGAATCG 29078
QY 571 ATCTGCTGAAGGCGCAACGATATCATCTGCGCGCTTGGCTTGGCC---TGGCGATCC 627
D 29079 ATTCTGCTCAAGCACAGATCTTGCCTTGCACATAGTTTCGCGCGGAGGATCGA 29138
QY 628 GACGAGCTTACCGT---TCCGAGGTTGCTCGGCGCAAGCCAAAGTGTTTACGATCTG 684
D 29139 AAGCAACAACTCGTGTGTTGACGATGACGCTCGGAGCAAGGCTTTCAGATCTT 29198
QY 685 CTTCGCGAGCATTCGGTCCGCTTTTCGAGGCGATGATCGAAGCAACCTGCAATCCAT 744
D 29199 ATTGCGGACATGGGTCCCGAGTGTGAGGAACTATCGGGGCGAAGTGTCCCGCAA 29258
QY 745 GCGAATCTTGCAGCAAGCGCTCGCCATCAGCATATCGATCTGGCTGCGGCGGCTACTC 804
D 29259 GGTGCTACGCGGAAATAATGTTAGCGAATATCTCCATTTGGCTCCCGGCTTCTC 29318
QY 805 AAGTCTCGAAGCGTGGCGGATCCCGAGTTCAGGATTCGATGATGATGATGATGATGATGAT 864
D 29319 AAGTCTCGAAGCGTGGCGGATCCCGAGTTCAGGATTCGATGATGATGATGATGATGATGAT 29378
QY 865 GAGACCGACCTCTACTTCCAGACGCTGGGCAAGTCTGAGCTCAAAAGGAAGCGCA 924
D 29379 GAAACACACACTATTACTTCCAACTCTTGGCAACCATGTGCAATGACGAGGAACG 29438
QY 925 GACTCTTCGAGCGAGATTCACGAAATGGTGGCTGCGGCTGCGGCTTACGGCTTCAAT 984
D 29439 AAGAAATACGAAACAGGATTCGAAGCAAGTGGAAACCGGATGCGGCTCGAAGGATTCAC 29498
QY 985 GATCAGCAGATCATGCGGACGATGATCATGAGCGCTTCTACGCTGATGATGCGGCTTGG 1044
D 29499 AACGATGACATCTGGCTCGGAGCTATGATGATGATGATGATGATGATGATGATGATGAT 29558
QY 1045 TCCGAAGAAATCTGTTGCGCGCGGACCGCGCAATCATCGAGTGGCGGCGCTTGGCAGT 1104
D 29559 GTCAACGAGATTTGTTGAGGTGAGCGAGGCTATCGTGGCATGCGGCAAGCTGCGGAGC 29618
QY 1105 CAGCAGATTCGCGGATTCAG 1125
D 29619 GAACACATCAGGATTCAG 29639

RESULT 10
AB088420 199035 bp DNA circular BCT 03-MAY-2003
LOCUS AB088420
DEFINITION Pseudomonas resinovorans plasmid pCAR1, complete sequence.
ACCESSION AB088420
VERSION AB088420.1 GI:26106080
KEYWORDS Pseudomonas resinovorans
SOURCE Pseudomonas resinovorans
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1
AUTHORS Maeda,K., Nojiri,H., Shintani,M., Yoshida,T., Habe,H. and Omori,T.
TITLE Complete nucleotide sequence of carbazole/dioxin-degrading plasmid
pCAR1 in Pseudomonas resinovorans strain CA10 indicates its
```

```
JOURNAL J. Mol. Biol. 326 (1), 21-33 (2003)
MEDLINE 22435905
PUBMED 12547188
REFERENCE 2 (bases 1 to 199035)
AUTHORS Nojiri,H., Shintani,M. and Maeda,K.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-2002) Hideaki Nojiri, Biotechnology Research
Center, The University of Tokyo; 1-1-1 Yayoi, Bunkyo-ku, Tokyo
113-8657, Japan [E-mail:anojiri@mail.ecc.u-tokyo.ac.jp,
Tel:81-3-5841-3064, Fax:81-3-5841-8030]
FEATURES
Location/Qualifiers
1..199035
/organism="Pseudomonas resinovorans"
/mol_type="genomic DNA"
/db_xref="taxon:53412"
/plasmid="pCAR1"
join(185744..199035,1..59548)
repeat_region
transposon="Tn4676"
gene
complement(join(198367..199035,1..81))
CDS
/gene="ORF188"
complement(join(198367..199035,1..81))
/function="putative ABC-transporter subunit"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="putative ABC-transporter subunit"
/protein_id="BAC41710.1"
/db_xref="GI:26106270"
/translation="MNPLSVQALEKTFGAVKATNGASFDIAPGELHAIIGPNAGKGS
TLIAOAGEIRPDKGRIFLNGODITDPAHORPLRGLARFQVQLYFEFLLENVAV
AIAADSGTFLKPMASDQSLAPADYLRVLEARGHAQVSALSGEROLEIAL
ALAQPSVLLDEPWAGNAGQESARMTQLLELGRYALLIVEHDMADVAFALADRTV
LVYGTILTSVEQNDPNPRAAYLGEHA"
gene
complement(1108..1929)
CDS
/gene="ORF31"
complement(1108..1929)
/function="putative membrane spanning subunit"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="putative ABC transporter subunit"
/protein_id="BAC41521.1"
/db_xref="GI:26106081"
/translation="MGVTLNLAGFYMVGAFCAAVTASLGSFWMAGLAAYIGSALYV
LLIETSNRRLYNRDLHQVLATLAVFTNELITVLGRSPMPPTWTSFVFL
PGLMTPVSRLLFIAAGAVATAMLLLNHRLMIRAGADHENVGALGNARLYT
LVFVFCVLGLAGFMAAPLISVEIGMERKVLITTFVIVVGGVSGVAGALLIG
MYDGLGRAYIPQLDQLLPABVSGTSLGGLSASAYIVMAVLLVVRPEGLLPARA"
complement(2084..3244)
gene
/gene="ORF30"
complement(2084..3244)
CDS
/gene="ORF30"
complement(2084..3244)
/evidence=not experimental
/codon_start=1
/transl_table=1
/product="putative ABC transporter subunit"
/protein_id="BAC41522.1"
/db_xref="GI:26106082"
/translation="MKKQLVGTLLAASCLASLAACQDPVRIQFTTTLTSGVLGSD
SRDGFQLADQEGKLGVPVSLNVEDDGLQAKKQIIIDMSLDGLISLYTGVFSNV
LAAVNTATKQGYIISNTGPSLQKQKPHRYFVASYQNDITHMAGVAANDLGYKK
WVILAPNTQGGDALDEGFKTFKEVETITKLNQDLDFSELRVSLAPDAIFQFHP
GGAGINFARQYSGSLNKSIPWVPVFMDEMLAATGDVAKGNNIVTLWPGSTNPA
NGAFVAKNDKVRPTVYAQTVDARLIGALKTVQGNLKDPEALRSALRNVTFDS
NGDFAFGNKNOHVIDWLLRVEADANGKLVQVPVQTIAKQVDSFAEAL"
complement(3255..3258)
repeat_unit
/rpt_type=direct
repeat_region
3259..4448
/insertion_seq="ISPre2"
repeat_unit
3259..3274
```

```
JOURNAL J. Mol. Biol. 326 (1), 21-33 (2003)
MEDLINE 22435905
PUBMED 12547188
REFERENCE 2 (bases 1 to 199035)
AUTHORS Nojiri,H., Shintani,M. and Maeda,K.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-2002) Hideaki Nojiri, Biotechnology Research
Center, The University of Tokyo; 1-1-1 Yayoi, Bunkyo-ku, Tokyo
113-8657, Japan [E-mail:anojiri@mail.ecc.u-tokyo.ac.jp,
Tel:81-3-5841-3064, Fax:81-3-5841-8030]
FEATURES
Location/Qualifiers
1..199035
/organism="Pseudomonas resinovorans"
/mol_type="genomic DNA"
/db_xref="taxon:53412"
/plasmid="pCAR1"
join(185744..199035,1..59548)
repeat_region
transposon="Tn4676"
gene
complement(join(198367..199035,1..81))
CDS
/gene="ORF188"
complement(join(198367..199035,1..81))
/function="putative ABC-transporter subunit"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="putative ABC-transporter subunit"
/protein_id="BAC41710.1"
/db_xref="GI:26106270"
/translation="MNPLSVQALEKTFGAVKATNGASFDIAPGELHAIIGPNAGKGS
TLIAOAGEIRPDKGRIFLNGODITDPAHORPLRGLARFQVQLYFEFLLENVAV
AIAADSGTFLKPMASDQSLAPADYLRVLEARGHAQVSALSGEROLEIAL
ALAQPSVLLDEPWAGNAGQESARMTQLLELGRYALLIVEHDMADVAFALADRTV
LVYGTILTSVEQNDPNPRAAYLGEHA"
gene
complement(1108..1929)
CDS
/gene="ORF31"
complement(1108..1929)
/function="putative membrane spanning subunit"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="putative ABC transporter subunit"
/protein_id="BAC41521.1"
/db_xref="GI:26106081"
/translation="MGVTLNLAGFYMVGAFCAAVTASLGSFWMAGLAAYIGSALYV
LLIETSNRRLYNRDLHQVLATLAVFTNELITVLGRSPMPPTWTSFVFL
PGLMTPVSRLLFIAAGAVATAMLLLNHRLMIRAGADHENVGALGNARLYT
LVFVFCVLGLAGFMAAPLISVEIGMERKVLITTFVIVVGGVSGVAGALLIG
MYDGLGRAYIPQLDQLLPABVSGTSLGGLSASAYIVMAVLLVVRPEGLLPARA"
complement(2084..3244)
gene
/gene="ORF30"
complement(2084..3244)
CDS
/gene="ORF30"
complement(2084..3244)
/evidence=not experimental
/codon_start=1
/transl_table=1
/product="putative ABC transporter subunit"
/protein_id="BAC41522.1"
/db_xref="GI:26106082"
/translation="MKKQLVGTLLAASCLASLAACQDPVRIQFTTTLTSGVLGSD
SRDGFQLADQEGKLGVPVSLNVEDDGLQAKKQIIIDMSLDGLISLYTGVFSNV
LAAVNTATKQGYIISNTGPSLQKQKPHRYFVASYQNDITHMAGVAANDLGYKK
WVILAPNTQGGDALDEGFKTFKEVETITKLNQDLDFSELRVSLAPDAIFQFHP
GGAGINFARQYSGSLNKSIPWVPVFMDEMLAATGDVAKGNNIVTLWPGSTNPA
NGAFVAKNDKVRPTVYAQTVDARLIGALKTVQGNLKDPEALRSALRNVTFDS
NGDFAFGNKNOHVIDWLLRVEADANGKLVQVPVQTIAKQVDSFAEAL"
complement(3255..3258)
repeat_unit
/rpt_type=direct
repeat_region
3259..4448
/insertion_seq="ISPre2"
repeat_unit
3259..3274
```

```

/rpt_type=inverted
complement (3404..4384)
/gene="tnpA3"
complement (3404..4384)
/note="transposase of ISPre2"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC41523.1"
/db_xref="GI:26106083"
/translation="MKOMTFADAEYAGKQKTRKELFLIEMDRVVPWKGLIALIEPHY.
PKGEGRPAYPLMAMLRVHLQNWFGYSDPAMEEALYETTLRQFAGLSLERIPDET
ILNFRLLERHELAAGILAVINGYLGDRGLSRQGTIVDATLINAPSTKNKQGRDP
EMHQTCKGNYIFGMKAHIGVDDSEGLVSHVGTAAANVADVTQVDKLHGDENVGAD
AGYTVKEPKEHEGREVINQIAARSTYKLDKRSALYKAARKIEKAQAQVRAKVEHP
FVTKRQFGYTKVRFRGLAKNTAQLVTLFALSNLMMARRHLLTNAGEVRL"
repeat_unit 4433..4448
/rpt_type=inverted
4444..4447
/rpt_type=direct
complement (4543..5568)
/gene="antC"
complement (4543..5568)
/gene="antC"
/note="AntC"
/codon_start=1
/transl_table=11
/product="reductase component of anthranilate"
/protein_id="BAC41524.1"
/db_xref="GI:26106084"
/translation="MSHKVAFSPADGKTLFFPVQSHIELLDAALENGINILDCREGV
CGTCQKCSQSQYDDEALAEEDLQKRMKTCQTRVSSDAAPYDFASLSCNAP
GPRLGIVRAVEQSPNTAILHLDAGQQLDFQGYARLQIPGTHGQRSYSFAN
RSGNQGLFLRLPDGLMSNYIERSLIGDEILAEPLGAFYLRHVDPKPLVFAVG
TGSLAFLMDLAEKGGCHPVHLYGVNNAADLCETORIAAVERIPGFFTPVIS
DRPDWHKRGYITTEHFDLAEFRDQALDMLVCGPPMVESIQMLDQTALDQTRYIE
KPTQSN"
complement (5587..6078)
/gene="antB"
complement (5587..6078)
/gene="antB"
/note="AntB"
/codon_start=1
/transl_table=11
/product="small subunit of terminal oxygenase component of
anthranilate 1,2-dioxygenase"
/protein_id="BAC41525.1"
/db_xref="GI:26106085"
/translation="MNSQLQYRVEQFLYRNABACDAQDWDYLSMFSDEPHLPQWD
SEHEYTRDPKRAMSLIYPNRGLEDVRVIRTKSASTVPMPTLHLLSNVRIAPLA
DDELEVQVNWHTLYRLQTAEQFFGRTYLRPDGSEWIKARKHVLNLDNTINSVLDF
YHL"
complement (6081..7484)
/gene="antA"
complement (6081..7484)
/gene="antA"
/note="AntA"
/codon_start=1
/transl_table=11
/product="large subunit of terminal oxygenase component of
anthranilate 1,2-dioxygenase"
/protein_id="BAC41526.1"
/db_xref="GI:26106086"
/translation="MSSARVSEOWKNFIEGCLDFPADGVFRIARDIFTEPDLFPLEM
ELIFENKMLYACHESIANKHDMWRAGROPMLITRDEGOLNALINACOHGTTLT
RVGKNQSTFTCFPHAWCYKSDRLVKNVAPGEYEGFPKATRGKAKRIESYKGFVP
ISLDVQGLSQFLGDAKVPFDMVAQSTGTEVLPGKSAITTDGNWKLQENGLD
GHVSTVHYNATVQROOVNSEKGLTSADLDYKLGAGDKETDDGFAPFNHGSV
LFDMPNPSVPGYATIMPRLQEQYQCGKAQWMMHRLNIIYPSMFFLDQISSQLRI
IRFIANNKTEIISFCVLGVGSEADRENRIQFDFPNVSGMGTDPDLVEFFEAQRGF
QARLERSDISRGVKGWVGPTANSETIGISPLVTETETHEGLYVNHQGNWRFLLE
GLARKAEQGPLKREV"

```

```

repeat_unit 7676..7679
/rpt_type=direct
7680..8871
/insertion_seq="ISPre1"
7686..7695
/rpt_type=inverted
complement (7826..8806)
/gene="tnpA2"
complement (7826..8806)
/gene="tnpA2"
/note="transposase of ISPre1"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC41527.1"
/db_xref="GI:26106087"
/translation="MKOMTFADAEYAGKQKTRKELFLIEMDRVVPWKGLIALIEPHY
PKGEGRPAYPLMAMLRVHLQNWFGYSDPAMEEALYETTLRQFAGLSLERIPDET
ILNFRLLERHELAAGILAVINGYLGDRGLSRQGTIVDATLINAPSTKNKQGRDP
EMHQTCKGNYIFGMKAHIGVDDSEGLVSHVGTAAANVADVTQVDKLHGDENVGAD
AGYTVKEPKEHEGREVINQIAARSTYKLDKRSALYKAARKIEKAQAQVRAKVEHP
FVTKRQFGYTKVRFRGLAKNTAQLVTLFALSNLMMARRHLLTNAGEVRL"
8856..8871
/rpt_type=inverted
8872..8875
/rpt_type=direct
9059..10315
/gene="ORF24"
9059..10315
/gene="ORF24"
/function="putative channel-forming protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Phak-like protein"
/protein_id="BAC41528.1"
/db_xref="GI:26106088"
/translation="MFOASHFTSMOTASLAALMTPLAQADFVEDSOLALGLNRFYI
DRFDGDNPNXSRVGSWTQGFDLRFNSGYTEGLQFALDASQVAYRLDGGGGRPS
IIPXTRRQOATIDYGRGGYAKRYSIKELTYGHRPTLPVATIDDSRQLVITYHGT
LIESKVDKLTLTGGRFTEISGRESSNSENKXVLFNGPDVORVDRGLNFGASVAFPTA
Query Match 36.9%; Score 419.4; DB 1; Length 199035;
Best Local Similarity 62.6%; Pred. No. 1.8e-85;
Matches 589; Conservative 0; Mismatches 406; Indels 6; Gaps 2;
QY 31 AGAACCAAGTTTGGGAGCCTTATATCCGTGCGAAACTCGGTTCCGAAACCATTTGGTAT 90
Db 28517 AGAGTAAAGGCTGGGCGCCCTACGTGGATGCGAAGCTAGGCTTTTCGCAATCATTTGGTAC 28576
QY 91 CCGGTTCGCTCGCGAGCGAAATCGCGAAGGTACTCCCGTTCCCGTCAAGCTCCTCGGA 150
Db 28577 CCGGTGATGTTTTCGAAGAGATCGACGAGGCGGCGAGACACTAATACTGCTCGGT 28636
QY 151 GAGAAGATTGCTCAATTCGCGTGGCGGCAAGGTCTATGCGATCCAGGACAGTGCCTG 210
Db 28637 GAGAACTTGTCTGCTCAATCGTATCGATGCGAAGCTGATTTGCCCTCAAGGACCGCTG 28696
QY 211 CATCGCGTGAACGCTTTCCGACCGGTCGAGTCTATTCCAGAGAACACCATATCTCGC 270
Db 28697 CATCGCGCGTCCAGTTGTTCGCTCAAAGTCAGTGCAGAACAGAGTGCACATGCATGC 28756
QY 271 TGGTATCAGCGCTGGACATATCGTGGGACGATGCGCGCTTCGTCGATATCTCTCACAAC 330
Db 28757 TGGTACCACGCTGGACCTATCGCTGGGAAGACGGCGTCTGTGTCGACATCTTGACCAAT 28816
QY 331 CCGGCGAGTGGAGATCGCGCGCGCGCTTTGAGACGTTTCCCGTTGAGAGGCCAAA 390
Db 28817 CCGAAGCGCACAGATCGGTGCGACAAAGCTGAAACTTACCCAGTGCAGAGGCCAAG 28876
QY 391 GGTCTTATCTTCGTTTACGTAGGCGACGGCGAAACCAACGGCGCTTATCGAAGATGTGCCG 450
Db 28877 GGCTCGCTTTCATTTCATTTCGTCGATGGCGACCTCTCTCTTGGCCCGGATACGCCA 28936

```

```
152-0035, Japan (E-mail:anojiri@mail.ecc.u-tokyo.ac.jp,
Tel:81-3-5841-3067, Fax:81-3-5841-8030)
Location/Qualifiers
1. .8096
/organism="Pseudomonas sp. K23"
/mol_type="genomic DNA"
/strain="K23"
/db_xref="taxon:202964"
1. .1113
/genes="ORF14K23L"
<1. .1113
/genes="ORF14K23L"
/function="unknown"
/notes="100% identity with the partial ORF14 in strain
CA10"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56720.1"
/db_xref="GI:28201183"
/translation="PQHGQACRQTPVQRPFPVWLPVSTIGVGVKSGAQRPP
LCVAVMTAFAYDPAADGAGSDLLNMQAQPAFYMALVDVALLGAAPFKAAARHH
WAPVNALANTPLQAFGEVAPHLIALGDTQERQAGTQALLTGTAGTALFLRSAYPA
QALQLFLYLAQVQGRARAVHLRFPADSRILPGLLOTLPQSQRRVADIVQDWRGCD
RQHGQVGLALHEAVPAOIIDTAPHLHLCPEOLAALLTQAEADAVFYOLTQHPPELLP
QOQRAFHORLATCLSTASEMNLSTQEDRSLFVLLSLNVEFYRCPALAPTWDVIGE
QGSLSQEQCGWSAETLQALMDFKASTVVDGNPL"
complement(1313..2341)
/genes="ORF14K23L"
complement(1313..2341)
/genes="ORF14K23L"
/function="transposase"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC56721.1"
/db_xref="GI:28201184"
/translation="NSYSELVSVERATIOVSHAQGLSLRRIACLNRSPTSVSREMR
NRDAQSVARVAQRMKARQACPMRKLPGSRFELVIMHRLRQLSPEOLAGKLR
SNIPSLRDVCRETINAYALPVGLKRLILCLRGKTRRPLGSDVRRGOIP
EMVSHLRPEVEDRLMPGHWDGLKIGKANASSVGLTVERTSGVLMVMDATATS
ANEGFSAALNNMPLMKRNTIDQGRMAQHABITORTGVAIYFCDPHSPWQSGNEN
INGLITKYLKFGTDLSVSHQEBELDAIAFQNLNMRPRKFRDFKCPIEVMSEVMOKALAQ
YDAPASIQ"
2760..3248
/genes="ORF12K23L"
2760..3248
/genes="ORF12K23L"
/function="unknown"
/notes="100% identity with ORF12 in strain CA10"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56722.1"
/db_xref="GI:28201185"
/translation="MNAFQIGRQHLRQLLSLVLAVLVGCKAEPDKGPATASISAY
NHTADYIHFFYVDGAWGNSRAYGGGVCCISYPRWRPDLTATVRWSTSSSPKPT
FSGCTWHELVPIEYDKLSLVHFLPNHQLVLLIWNWSAGSKGYRGDPDAPEKPADW
PY"
3274..3786
/genes="ORF11K23L"
3274..3786
/genes="ORF11K23L"
/function="unknown"
/notes="100% identity with ORF11 in strain CA10"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56723.1"
/db_xref="GI:28201186"
/translation="MNAFQIGRQHLRQLLSLVLAVLVGCKAEPDKGPATASISAY
NHTADYIHFFYVDGAWGNSRAYGGGVCCISYPRWRPDLTATVRWSTSSSPKPT
FSGCTWHELVPIEYDKLSLVHFLPNHQLVLLIWNWSAGSKGYRGDPDAPEKPADW
YGTGTTWHEKVPYIEYDKLSLVHFLPNHQLVLLIWNWSAGSKGYRGDPDAPEKPADW
PY"

FEATURES
source
152-0035, Japan (E-mail:anojiri@mail.ecc.u-tokyo.ac.jp,
Tel:81-3-5841-3067, Fax:81-3-5841-8030)
Location/Qualifiers
1. .8096
/organism="Pseudomonas sp. K23"
/mol_type="genomic DNA"
/strain="K23"
/db_xref="taxon:202964"
1. .1113
/genes="ORF14K23L"
<1. .1113
/genes="ORF14K23L"
/function="unknown"
/notes="100% identity with the partial ORF14 in strain
CA10"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56720.1"
/db_xref="GI:28201183"
/translation="PQHGQACRQTPVQRPFPVWLPVSTIGVGVKSGAQRPP
LCVAVMTAFAYDPAADGAGSDLLNMQAQPAFYMALVDVALLGAAPFKAAARHH
WAPVNALANTPLQAFGEVAPHLIALGDTQERQAGTQALLTGTAGTALFLRSAYPA
QALQLFLYLAQVQGRARAVHLRFPADSRILPGLLOTLPQSQRRVADIVQDWRGCD
RQHGQVGLALHEAVPAOIIDTAPHLHLCPEOLAALLTQAEADAVFYOLTQHPPELLP
QOQRAFHORLATCLSTASEMNLSTQEDRSLFVLLSLNVEFYRCPALAPTWDVIGE
QGSLSQEQCGWSAETLQALMDFKASTVVDGNPL"
complement(1313..2341)
/genes="ORF14K23L"
complement(1313..2341)
/genes="ORF14K23L"
/function="transposase"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC56721.1"
/db_xref="GI:28201184"
/translation="NSYSELVSVERATIOVSHAQGLSLRRIACLNRSPTSVSREMR
NRDAQSVARVAQRMKARQACPMRKLPGSRFELVIMHRLRQLSPEOLAGKLR
SNIPSLRDVCRETINAYALPVGLKRLILCLRGKTRRPLGSDVRRGOIP
EMVSHLRPEVEDRLMPGHWDGLKIGKANASSVGLTVERTSGVLMVMDATATS
ANEGFSAALNNMPLMKRNTIDQGRMAQHABITORTGVAIYFCDPHSPWQSGNEN
INGLITKYLKFGTDLSVSHQEBELDAIAFQNLNMRPRKFRDFKCPIEVMSEVMOKALAQ
YDAPASIQ"
2760..3248
/genes="ORF12K23L"
2760..3248
/genes="ORF12K23L"
/function="unknown"
/notes="100% identity with ORF12 in strain CA10"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56722.1"
/db_xref="GI:28201185"
/translation="MNAFQIGRQHLRQLLSLVLAVLVGCKAEPDKGPATASISAY
NHTADYIHFFYVDGAWGNSRAYGGGVCCISYPRWRPDLTATVRWSTSSSPKPT
FSGCTWHELVPIEYDKLSLVHFLPNHQLVLLIWNWSAGSKGYRGDPDAPEKPADW
PY"
3274..3786
/genes="ORF11K23L"
3274..3786
/genes="ORF11K23L"
/function="unknown"
/notes="100% identity with ORF11 in strain CA10"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56723.1"
/db_xref="GI:28201186"
/translation="MNAFQIGRQHLRQLLSLVLAVLVGCKAEPDKGPATASISAY
NHTADYIHFFYVDGAWGNSRAYGGGVCCISYPRWRPDLTATVRWSTSSSPKPT
FSGCTWHELVPIEYDKLSLVHFLPNHQLVLLIWNWSAGSKGYRGDPDAPEKPADW
YGTGTTWHEKVPYIEYDKLSLVHFLPNHQLVLLIWNWSAGSKGYRGDPDAPEKPADW
PY"

RESULTS
AB094786
LOCUS
DEFINITION
Pseudomonas sp. K23 car operon (ORF14K23L, ORF12K23L,
ORF11K23L, ORF10K23L, ORF9K23L, ORF4K23L), partial and complete
cds.
ACCESSION
AB094786
VERSION
AB094786.1 GI:28201182
KEYWORDS
Pseudomonas sp. K23
SOURCE
Pseudomonas sp. K23
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 Inoue, K. and Nojiri, H.
Divergent structures of carbazole-degrading car operons isolated
from several gram-negative bacteria
Unpublished
JOURNAL
2 (bases 1 to 8096)
AUTHORS
Nojiri, H. and Inoue, K.
TITLE
Direct Submission
JOURNAL
Submitted (28-OCT-2002) Hideaki Nojiri, Biotechnology Research
Center The University of Tokyo; Yayoi 1-1-1, Bunkyo-ku, Tokyo
```


gene	PKTYLEDLK"	Db	7444	GAGAACTTGTCTGTCATCGTATCGATGGAAGCTGATTGCTCAGGACCGCTGCTG	7503
CDS	4160..5140	Qy	211	CATCGCGGTGTAAAGCTTTCCGACCGGGTTCAGTGTCTATTCCAGAAACACATATCTGTC	270
	4160..5140	Db	7504	CATCGCGCGTCCAGTTTGTCTGCTCAAAAGTCTGCAAAACGAAAGTCTGACATCATGC	7563
	/function="transposase"	Qy	271	TGGTATCACCGCTGGACATATCGTGGAGAGTGGCGCTCTGTCGATATCTTCCACAAAC	330
	/notes="100% identity with TnpA1,TnpA2 and TnpA4 in strain Cal10"	Db	7564	TGGTATCACCGCTGGACATATCGTGGAGAGAGTGGCGCTCTGTCGATATCTTCCACAAAC	7623
	/codon start=1	Qy	331	CCCGCGAGTGTGACATCGCGCGCGCTTTGAAGACGTTTCCGGTGTGAAGAGGCGAAA	390
	/transl_table=11	Db	7624	CCGACAAAGCGCACAGATCGGTTCGACAAAAGCTGAAAACCTTACCAGTGCAGGAAGCAAG	7683
	/product="transposase"	Qy	391	GGTCTTATCTCTGTTTACGTAGCGCGGCAACCAACGCGCTTATCGAAGATGTGCGG	450
	/protein_id="BAC56724.1"	Db	7684	GGTCTTATCTCTGTTTACGTAGCGCGGCAACCAACGCGCTTATCGAAGATGTGCGG	7743
	/db_xref="GI:28201187"	Qy	451	CCCCGGCTTCTTGTATGAAAACCGCGCATTCACGGCCCAACATCGGCTCGTGGCTCGAAC	510
gene	5508..7211	Db	7744	CCCAATTTCTTGACGATGACATGGAATCCTCGGGAAGAACCAATCATCAAGTCTAAC	7803
CDS	5508..7211	Qy	511	TGGCGCTTGGTGGGAAAAGCGCTTTGATCGGGGCAACGCTCTTCAATTCAGAGAATCG	570
	/genes="ORF9K23L"	Db	7804	TGGCGCTTGGTGGGAAAAGCGCTTTGATCGGGGCAACGCTCTTCAATTCAGAGAATCG	7863
	/function="unknown"	Qy	571	ATCTGTGTGAAGGCAACGATATCATCTGCGCTTGGCTTTTGGCGC---TGGCGATCCC	627
	/notes="100% identity with ORF9 in strain Cal10"	Db	7864	ATCTGTGTGAAGGCAACGATATCATCTGCGCTTGGCTTTTGGCGC---TGGCGATCCC	7923
	/codon start=1	Qy	628	GACGAGTTAGCGT---TCCGAGTTTCTGCGGGCAAGCCCAAGGTTTACGATCTG	684
	/transl_table=11	Db	7924	AAGCAACAACTCGTGTGTTGACATGACGTCGTCGGAGCGCAAGGTTTACGATCTT	7983
	/product="partial terminal oxygenase component of carboxole"	Qy	685	CTTCGCGAGATTCGGTGGCGGTTTTCGAAAGGATGATCGAAGCAACCTGCAATCAT	744
	/protein_id="BAC56726.1"	Db	7984	ATTGCGCAACATGGGGTCCCAAGTCTTTCAGGAACTATCGGGGCGAAGTGGTCCGCGAA	8043
	/db_xref="GI:28201189"	Qy	745	GGCAACATTCGACAGCAAGCGCTGCCATCAGCATATCGATTCGATTCGCTGGCGGG	797
gene	7294..8096	Db	8044	GGTCCCTACGGCGGAAAATTTAGCGAAGCATATCTCCATTTGGCTCCCGGG	8096
CDS	7294..8096	RESULT 12			
	/genes="ORF4K23L"	PPYL2655			
	/notes="100% identity with partial CarAa in strain Cal10"	LOCUS			
	/codon start=1	DEFINITION			
	/transl_table=11	P.putida oxoS, oxoO, oxoH genes.			
	/product="partial terminal oxygenase component of carboxole"	ACCESSION			
	/protein_id="BAC56726.1"	Y12655			
	/db_xref="GI:28201189"	Y12655.1			
		GI:2072730			
		2-oxo-1,2-dihydroquinoline 8-monooxygenase; oxoH gene; oxoO gene; protein.			
		oxoS gene; oxygenase; putative hydrolase; putative regulatory			
		protein.			
		Pseudomonas putida			
		Pseudomonas putida			
		Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;			
		Pseudomonadaceae; Pseudomonas.			
		1			
		Rosche, B., Tahisuka, B., Hauer, B., Linsens, F. and Fetzner, S.			
		2-oxo-1,2-dihydroquinoline 8-monooxygenase: phylogenetic			
		relationship to other multicomponent norheme iron oxygenases			
		J. Bacteriol. 179 (11), 3549-3554 (1997)			
		97315225			
		9171399			
		2 (bases 1 to 3203)			
		Fetzner, S.			
		Direct Submission			
		Submitted (17-APR-1997) S. Fetzner, Universitaet Oldenburg,			
		Fachbereich 7, Postfach 2503, D-26111, Oldenburg, FRG			
		Location/Qualifiers			
		1. 3203			
		/organism="Pseudomonas putida"			
		/mol_types="genomic DNA"			
		/strain="86"			
		source			
		FEATURES			
		ORIGIN			
		Query Match			
		Best Local Similarity			
		Matches			
		24.8%; Score 281.8; DB 1; Length 8096;			
		62.1%; Pred. No. 6.4e-54;			
		Conservative 0; Mismatches 287; Indels 6; Gaps 2;			
Qy	31	AGAACCAAGTTTGGAGCGCTTATATCCGTGCGAAACTCGGTTCCGAAACCATGGTAT	90		
Db	7324	AGAGTAAAGGCTGGGCGCCTACGTGGATGCGAAGCTAGGCTTCCGAATCATTTGGTAC	7383		
Qy	91	CCCGTTCCGCTCGGAGCGAAATCGCGAAGTACTCCCGTTCCCGCTCAAGTCTCTGGGA	150		
Db	7384	CCGGTATGTTTTCGAAGAGATTCGACGAGGCGCGGACACATTAACCTGCTCGT	7443		
Qy	151	GAGAAGATTCTGCTCAATCGCTGGCGGCAAGGTTCTATCGCATCCAGACAGGTCGCTG	210		


```
gene
complement(1000..1377)
/gene="SC00858"
/notes="synonym: SCM2.11c"
CDS
complement(1000..1377)
/gene="SC00858"
/notes="SCM2.11c, unknown, len: 125 aa."
/codon_start=1
/transl_table=11
/product="hypothetical protein SCM2.11c."
/protein_id="CAB65639.1"
/db_xref="GI:6689149"
/db_xref="SPTREMBL:Q9RCV3"
/translation="MQEEFARSALDMPISAFNASSDDTYTALLSQALTSDDVVFWSPLG
RSGIAEVRFDIRKHPAGTGMVRCSAVDPEWARIRWVFSPDGGPLAGTVDV
VHLRRSLIDQVIFVAGDIEPAAS"
complement(1546..2055)
/gene="SC00859"
/notes="synonym: SCM2.12c"
complement(1546..2055)
/gene="SC00859"
/notes="SCM2.12c, possible membrane protein, len: 169 aa.
Contains possible hydrophobic membrane spanning region"
/codon_start=1
/transl_table=11
/product="putative membrane protein."
/protein_id="CAB65640.1"
/db_xref="GI:6689150"
/db_xref="SPTREMBL:Q9RCV2"
/translation="MTTSWTFPPRLADTALAALAPICACVLGALGAVWATGNAG
TPAIGVTDARLPLRGVGPETAFFKINTNGAQRDLVGVTSSEVPEGISSHRMT
AGGAHRRPTESLPVPAEGTLDMSPLSSDVTVPAAARWQAGDLVPPTLHFSGRMEV
LAVVRPGS"
complement(2052..4340)
/gene="SC00860"
/notes="synonym: SCM2.13c"
complement(2052..4340)
/gene="SC00860"
/notes="SCM2.13c, probable cation-transporting ATPase, len:
762 aa. Highly similar to many other cation transporters
e.g. Enterococcus hirae SW:COPA_ENTHR(EMBL:L13292)
coper/potassium-transporting ATPase A (EC 3.6.1.36) (727
aa), fasta scores opt: 1287 z-score: 1275.2 E(): 0 37.8%
identity in 749 aa overlap. Contains Prosite hits to
PS01047 Heavy-metal-associated domain and PS00154 E1-E2
ATPases phosphorylation site. Also contains Pfam matches
to entries PF00122 E1-E2 ATPase, E1-E2 ATPase and PF00403
HMA, Heavy-metal-associated domain as well as multiple
possible membrane spanning hydrophobic domains."
/codon_start=1
/transl_table=11
/product="probable cation-transporting ATPase"
/protein_id="CAB65641.1"
/db_xref="GI:6689151"
/db_xref="GOA:Q9RCV1"
/db_xref="SPTREMBL:Q9RCV1"
/translation="MAGEPSTSEVDTLAVGMTCACTVTRVERKLAKLDVGSVNL
ATGARVHPPEVLPOLVAASVQAGTALPOSVEERRGSDDDAGTQAQRDLRT
TVALLAVLVLSVLPQAWQPNWCLFVLAAPVAVGAWPFHRAARLRHSTMD
TLVSLGVASFWSSYALFTGGAGDPLDFPFLVPTASDGVHYLEAAGVPLFVL
AGRYLARARRGGAALRALAEVAVKVEAVRDGTGGRIPIEELRVGVQVVRPGERV
ATDGTVVGSSAVDLSLVTCSEPAEAVPGTAVIAGVNVGGLAVRATVAGADTFLA
RITHLVEAQAGKARQRLAKVGVFVPLVPLATLVGLFGLGAGADPQAALTASVA
VLVAVPCALGLATPTALMAATGRGAQLGVLVSGPQALSGLRHIDVVDLTKTGLTFSG
HMSVARTNPRFGIEQVAVLAGVEQSEHPLGQAVTAYARNTPTGSLPEVTDFA
ALPGRVGRVEGRLEVTAPDDELFPVLDPEAMSGAESAAHTPVVVRVDGTEGLIEV
RELREGRVAVVGDGVNDAAALAGADLGTAMGTGTDAAIGAADVTLVRGDDIALADA
VRLSRSTGLRVNLLWAFQNVVVTMPVAVVGLSPMPAAVMSVSSLLVGVNSLRLR
AWOPARTSPSPASRGESLR"
complement(2133..3707)
/gene="SC00860"
/notes="Pfam match to entry PF00122 E1-E2 ATPase, E1-E2
ATPase, score 529.80, E-value 1.9e-155"
misc_feature

misc_feature
complement(2997..3017)
/gene="SC00860"
/notes="PS00154 E1-E2 ATPases phosphorylation site"
complement(4209..4295)
/gene="SC00860"
/notes="Pfam match to entry PF00403 HMA,
Heavy-metal-associated domain, score 37.40, E-value
1.1e-08"
complement(4212..4295)
/gene="SC00860"
/notes="PS01047 Heavy-metal-associated domain"
complement(4340..4549)
/gene="SC00861"
/notes="synonym: SCM2.14c"
complement(4340..4549)
/gene="SC00861"
/notes="SCM2.14c, putative secreted protein, len: 69 aa.
Rich in the amino acid Gly. Contains possible N-terminal
region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAB65642.1"
/db_xref="GI:6689152"
/db_xref="SPTREMBL:Q9RCV0"
/translation="MQRIILPPLPCGFLVLLVLAFAASYAVGRGVGPVAFQMGPGGIT
QGGHGGDGDMEEDMGMMHGGGH"
complement(4813..5472)
/gene="SC00862"
complement(4813..5472)
/notes="synonym: SCM2.15c"
/gene="SC00862"
/notes="SCM2.15c, possible integral membrane protein, len:
219 aa. Similar to a protein of unknown function from
Mycobacterium tuberculosis SW:Y970 MYCTU(EMBL:Z79700)
hypothetical 22.9 KD protein (210 aa), fasta scores opt:
224 z-score: 235.4 E(): 9.8e-06 30.5% identity in 213 aa
overlap. Contains possible membrane spanning hydrophobic
domains and a possible N-terminal signal sequence."
/codon_start=1
/transl_table=11
/product="putative integral membrane protein."
/protein_id="CAB65643.1"
/db_xref="GI:6689153"
/db_xref="GOA:Q9RCU9"
/db_xref="SPTREMBL:Q9RCU9"
/translation="WIAAGLRWILLMEAPALYGLWRLVPTTGLTGRVGHLLHAA
NSVLMIAWNPWGMNDLAVAPQVLPFAAGALMFVAASLVWFGERSRSGAVRAAWPHALM
MGAWMVAAWGTSGAWAGHGGGTGGHGHGAASGSLSLTATGPSVASALLAV
ALTAIGLVMLARALDLARVAPLPAGGPAPAGTDTTAAALDPACHAVMALGMVAMFALF
A"
complement(5469..6926)
/gene="SC00863"
/notes="synonym: SCM2.16c"
complement(5469..6926)
/gene="SC00863"

Query Match 4.3%; Score 48.8; DB 1; Length 298450;
Best Local Similarity 48.2%; Pred. No. 1.5;
Matches 137; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 721 ATCGAGCGCAACCTGCATCCATCGCACATTGGCAGCAGCGCGTCCCATCAGGATA 780
DB 97961 ATGGACGACGATAGCGCTTCCCTGGTACTCAACACCGCGCCCTCTTCTCAACAAG 97902
QY 781 TCGATCTGCCTGCCGGGGGTACTCAAGGTCCGACCGTGGCCGGTCCCGAGCTCACGCG 840
DB 97901 TCCCTCTTCGAGAGGCCGGACTCGACCCGAGCAGCGCCGAGGACGTACGACGAGTC 97842
QY 841 TTCGAATGTGTAGTGCCTCGGTGATGACGACGACCTCTTCTTCCAGACCTCGGCHAA 900
DB 97841 TTCGACGCGCCCTGAAGATCGCGACCAAGACCGCGCGAGGTCCGCCACCTCGCCAAC 97782
```

QY 901 GTCTGTGACGTCACAAAGGAGCGGACAGATCCCTTCGAGCGGAGAAATCCAGAAAAATGGGTA 960
Db 97781 GTGCCACCATCGAGGACTTCGGCGCTACGGCGTCCGCTCATGAACAGGAGGAC 97722
QY 961 GGCCTTCGGCTTAACGGCTTCAATGATACACATCATGGCAG 1004
Db 97721 GCCTTCGGCTTCAACGAGCGCAAGGGTGTCCAGCTCCTCGCCG 97678

RESULT 15
AF429315 125020 bp DNA linear PRI 18-JAN-2002
LOCUS AF429315 Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 125020)
Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S.,
Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A.,
Potter, N.T., Ross, C.A. and Margolis, R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
Nat. Genet. 29 (4), 377-378 (2001)
1. 125020
2 (bases 1 to 125020)
Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.
Direct Submission
Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
Location/Qualifiers
1. 125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
complement(36507..36887)
/gene="JPH3"
/note="synonym: JP3"
complement(36507..36887)
/gene="JPH3"
/product="junctophilin 3"
complement(36507..36887)
/gene="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/protein_id="AA140941.1"
/db_xref="GI:17646245"
/translation="MSSSGRPNFDGSGYCGWEDGKAKHGVCCTGPKGQGEYTGWS
HGFEVLGYTPSGNYGTGWAQKRGHIGLESKGMVYKGEWTHGPKRGYVRECA
NGAKYEGTWSNGLQDGYGTETYSQ"

ORIGIN

Query Match 4.2%; Score 47.4; DB 9; Length 125020;
Best Local Similarity 11.5%; Pred. No. 3.1;
Matches 106; Conservative 374; Mismatches 439; Indels 5; Gaps 2;
QY 80 ACCATGTGATCCCGTTCGCTCGAGCGAAATCCGCAAGGTACTCCCGTCCCGTCA 139
Db 16929 NSMYTKSSWGSSGWCWRRRRSGKWKYSGRMSKSMRYTGGSKMNRSSMCTSSCY 16988

QY 140 AGCTCTGGGAGAGAGATTTCTCTCAATCGCGTGGCGGCAAGGTCTATCGATCAAG 199
Db 16989 ASWCCNCWSCCWRSCCCCMRSYCCMEYCCACKCYNMSYWTMSASYNRSRYSKSMK 17048
QY 200 ACAGGTGCTGCATCGCGGTGTAAAGCTTTCCGACCGGTGCGAGTCTATTCCAAGAAC 259
Db 17049 CWSRMCSSRSRSGKSGCGSGMGKSGKSGYGRKTKRSKRGKMGKWKYMYRRSRMRK 17108
QY 260 CCATATCTCTGCTGTATACAGGCTGACATATCGGTGGAGCGATGGCCCGCTCGTCA 319
Db 17109 KWSYSSKGMCMYCWGRRGCTCSCMTSRANCCSYCYAKCKSMCYCYGKSMKGYVA 17168
QY 320 TCCTCAAAACCCCGGCGAGTGTGAGATCGCGCGCGCGCTTTGAAGAGCTTCCCGGTG 379
Db 17169 CSYRSGMSKCYMRGSGTSTSCGCTTTTCCCGCANTGGGGAAGCTTTTNCNKKTY 17228
QY 380 AAGAGCCCAAGGCTTATCTCTGTTAGTGGAGCGCGCAACCAACGCGCTTATCG 439
Db 17229 YRKRNGCAMKYNYNNSWRSSCPAGMSTCYKSSMTMSVASYCWNMSYKSMRMA 17288
QY 440 AAGATGTGCGCGCGCTTCTCTGTATGAAAACCGGCCATTACCGGCCAACATCGGCTCG 499
Db 17289 SRGMSNSYMKMKSSWRMSCTMKCCMKMCMYCMRSMRSMRSMYMYASWKSRRGY 17348
QY 500 TGGCTTCGAACTGCGCTTGGGTGCGGAAACGGCTTTCATCGGGGCGACGCTTTCATC 559
Db 17349 CTRCYWC---MSKSKCYKSYMMRMRKMKWKSRCWMSGASRSCYKCYKSMRSC 17405
QY 560 ACAAGAAATTCGATCTGTTGAAGGCAACGATATCATCTCGCGCTTGGCTTGGCGCTG 619
Db 17406 MMSKCYRCAGCMWKGGMGYMYRMCWMMKRRWWSKSMYRMRMRWKGGMWCMKCYSR 17465
QY 620 GCGATCCGACAGCTTACGCTTCGAGGTTCCTCGGGGCAAGCCAAAGGTGTTAG 679
Db 17466 MFCMMMKSYCASCRSCAMSGTYNKCASYCMCMSSYWTCTCTSTYMYRCTCCK--G 17523
QY 680 ATCTGCTTGGCGAGCATTCGCTGCCGTTTTCGAGGCGATGATCGAAGGCAACCTGCAA 739
Db 17524 WSSYTKSKWSSSSMSYKGRKSYSMCTSRGAMSCWRRCCVMRGASSMRAGSMRRA 17583
QY 740 TCATGCGCAACATTCGACAGCGGTGCGCCATCAGCATATCGATCTGGCTGCCGCGG 799
Db 17584 KGRSWGGRSKWMTTGMWRSKYYYCTGRMMMTYMCWRRRSMYRSMAMGRKSS 17643
QY 800 TACTCAAGGTGCAACCGTGGCGGATCCCGAGTCCAGCGATTCGAATGATGATCGTCCG 859
Db 17644 WSGRMMGASRRCKSRSSWCSRMKMRGSCWSSKMWGSSRSRSCSKSGRGR 17703
QY 860 TCATGAGACCGCACCTCTACTTCAGACGCTGGGCAAGTCTGAGCTCAAGGAAG 919
Db 17704 RRSKRSKYRGRGRGRKRMSTKSGSKGSKCMKSGSMSTSSCYYSASSCMMMSKSC 17763
QY 920 CGGACAGCTCTTCGAGCGAGAAATTCACGAAAAATGGGTAGGCTCGCGCTTAACGGCT 979
Db 17764 MCCMMKRCACCYSMSTCTMYRCWGMGMSKSYSTCGTCTCKCTGKSRWYMYWY 17823
QY 980 TCAATGATGACGACATCATGGCAC 1003
Db 17824 TSNWYWRACWYCYMSTRMMMC 17847

Search completed: June 19, 2004, 06:08:07
Job time : 4598.53 secs